

❖ TWO ❖

AN EVALUATION THEORY TREE

Marvin C. Alkin and Christina A. Christie

Our evaluation theory tree is presented in Figure 2.1, in which we depict the trunk and the three primary branches of the family tree. The trunk is built on a dual foundation of accountability and systematic social inquiry. These two areas have supported the development of the field in different ways. The need and desire for accountability presents a need for evaluation. The importance of accounting for actions or for resources used in the conduct of programs is particularly evident for programs supported by government entities. The same accountability demand could be said to be present for corporate businesses that annually are required to provide reports from outside auditors to their shareholders (although the accounting scandals of the early 2000s, e.g., Enron, would dispute how accountable these audits actually have been). As a root for program evaluation, we think of accountability in the broadest way possible. That is, accountability is not a limiting activity, but, rather, is designed to improve and better programs and society. The social inquiry root of the tree emanates from a concern for employing a systematic and justifiable set of methods for determining accountability. While accountability provides the rationale, it is primarily from social inquiry that evaluation models have been derived.

The main branch of the tree is the continuation of the social inquiry trunk. This is the evaluation as research, or evaluation guided by research methods, branch. This branch we have designated *methods* since in its purest form, it

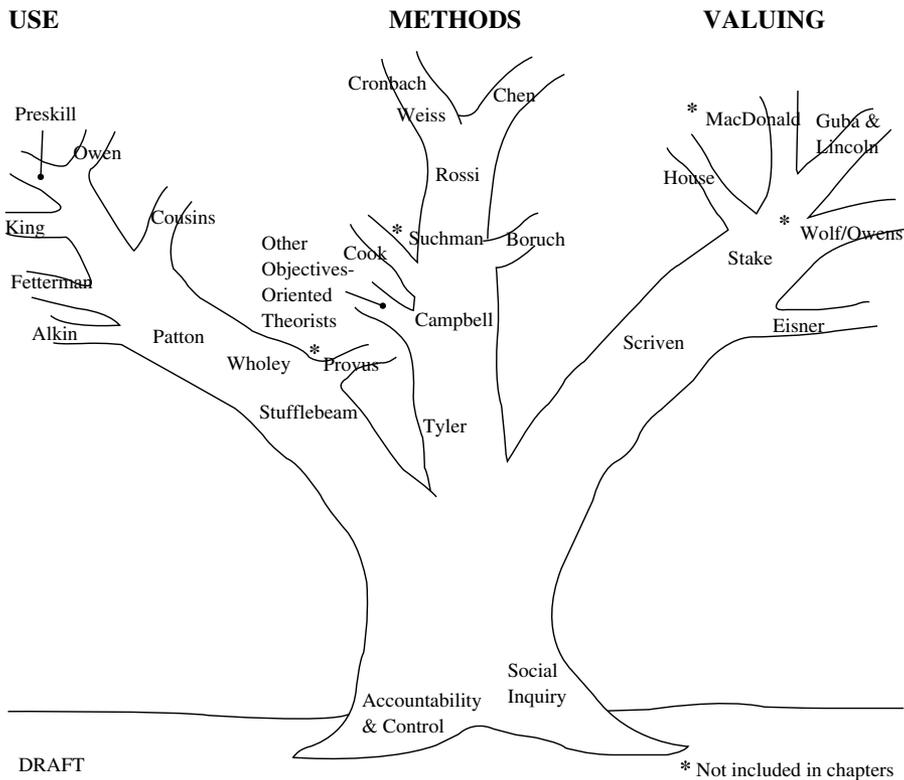


Figure 2.1 Evaluation Theory Tree

deals with obtaining generalizability, or “knowledge construction,” as Shadish, Cook, and Leviton (1991) refer to it. Another branch we call the *valu-*
ing branch. Initially inspired by the work of Michael Scriven (1967), the valu-
ing branch firmly establishes the vital role of the evaluator in valuing. Those
on this branch maintain that placing value on data is perhaps the most essen-
tial component of the evaluator’s work. Some subsequent theorists extend the
evaluator’s role to include systematically facilitating the placing of value by
others (e.g., Guba & Lincoln, 1989).

The third major branch is *use*, which, with the pioneering work of Daniel
Stufflebeam (initially with Egon Guba) and the work of Joseph Wholey, origi-
nally focused on an orientation toward evaluation and decision making. In
essence, work done by theorists on this branch expresses a concern for the way

in which evaluation information will be used and focuses on those who will use the information.

We will start with a discussion of the accountability and social inquiry roots and then discuss each of the major branches of the tree.

ACCOUNTABILITY AND CONTROL

Accountability refers to the process of “giving an account” or being answerable or capable of being accounted for. Wagner (1989) indicates that there are several dimensions to accountability. The first of these is “reporting,” in which description is provided. A second phase of accountability is a “justifying analysis” or explanation. But accountability in its fullest sense may be more. In instances where a justifying analysis recognizes deficiencies, true accountability requires “answerability”—that is, those responsible must be held accountable. This phase of accountability is not reflected in evaluation; evaluation simply provides the information for “being answerable.”

Alkin (1972a), in a paper defining accountability, refers to goal accountability, process accountability, and outcome accountability. Goal accountability examines whether reasonable and appropriate goals have been established. Governing boards and upper levels of management are the responsible entities for this kind of accountability. Process accountability reflects whether reasonable and appropriate procedures for accomplishing those goals have been established and implemented. Typically, management and program operators bear responsibility for process accountability. Outcome accountability refers to the extent to which established goals have been achieved. As in process accountability, management and program operators are to be held accountable for outcomes.

These three accountability types are reflected in the evaluation writings. Concern for the evaluator’s role in valuing goals is evident in Michael Scriven’s work. Program accountability is prominent in the “process” section of Daniel Stufflebeam’s CIPP model. Finally, outcome accountability, the provision of evaluation information for examining the adequacy of outcomes, is the major thrust of most evaluation efforts.

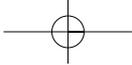
Because of its important historic role, process accountability perhaps requires further amplification. The tradition of process accountability, while less visible in today’s evaluation writing, nonetheless provides important roots

for evaluation. For example, the European tradition of school inspectors (now largely outdated) has been depicted in countless movies showing dreaded visits of the school inspector, whose presence is designed to ensure that teachers are engaged in performing prescribed lessons at prescribed times. Perhaps a contemporary evaluation procedure designed for obtaining process accountability is the system of school accreditation employed in secondary schools, private schools, universities, and professions. In these instances, evaluation procedures are designed whereby teams with presumed expertise, frequently guided by established process standards, visit a site to observe, account, and make a report. The result is a judgment about whether institutions are accountable and should be accredited.

Many theorists acknowledge the important role that accountability plays in evaluation. Mark, Henry, and Julnes (2000) cite it as one of the four purposes of evaluation. Today, most evaluations have a strong accountability thrust, having as a goal the improvement of institutional performance. The results of these evaluations are often used in policymaking or other governmental decision making. The work of the evaluation unit of the U.S. Government Accounting Office (GAO) typifies the kinds of accountability activities that reflect this broader scope. The GAO is independent, nonpartisan, and often described as the “congressional watchdog” because it investigates how the federal government spends tax dollars. The GAO’s mission is to “examine the use of public funds, evaluate federal programs and activities, and provide analyses, options, recommendations, and other assistance to help the Congress make effective oversight, policy, and funding decisions through financial audits and program reviews” (see www.gao.gov). It is important to note that this notion of accountability implies the provision of an account to one (or a group) in a position of formal authority. Thus, the GAO’s activities are designed to ensure the executive branch’s accountability to the Congress and the federal government’s accountability to all Americans.

SOCIAL INQUIRY

Social inquiry, in its broadest sense, can be characterized as the systemic study of the behavior of groups of individuals in various kinds of social settings by a variety of methods. It begins with the recognition that there is a unique social dimension to human action as opposed to merely a natural and psychological



dimension. The central overriding question is, Why do people in social groups act as they do? In the Western world, inquiry along these lines has its origins in 17th- and 18th-century figures such as Hobbes, Montesquieu, and Rousseau. While these theorists systematically studied and commented on social groups, their descriptions and theories were a product more of contemplation rather than rigorous empirical investigations. It wasn't until the mid- and late 19th century, as demonstrated in the works of Karl Marx, Emile Durkheim, and Max Weber, for instance, that society and social groups began to be studied empirically through the collection and analysis of empirical data on social groups.

A perennial question in social inquiry is which methods are appropriate for the study of society, social groups, and social life and whether the methodologies of the physical sciences, broadly defined, are applicable to social phenomena. The classical social theorists, Marx (1932) and Durkheim (1966), for example, made extensive use of statistics, amongst other data, in coming to form particular judgments regarding social life. The influence of the discipline of psychology introduced into the social realm the experimental method, where the central overriding question is whether a treatment is effective in bringing about desired effects. Discussions regarding the feasibility and desirability of this methodology for the study of the social world continues to this day, giving rise to heated debates and never-ending accusations of positivism. These debates are related to and reflective of the more general question of the applicability of the methods of the sciences to the social sciences.

Alternatively, the discipline of anthropology has given rise to ethnographies and, more broadly, qualitative studies of the social world. The distinction between these methods and the ones just mentioned above is sometimes couched in terms of the distinction between explanation and prediction, on one hand, and interpretation and understanding, on the other. Clifford Geertz's classical essay "Thick Description: Toward an Interpretive Theory of Culture" in *The Interpretation of Cultures* (1973) epitomizes the latter approach and in part has come to define interpretive social science where the emphasis is placed not on prediction, but on meaning.

Philosophers of science and social science continue to disagree on what constitutes the methods of the sciences and their potential applicability to the study of social life (Kuhn, 1962; Popper, 1963). Cutting across social science disciplines are broad philosophical and methodological questions that continue to be debated in contemporary social inquiry. Important questions



include: What is the relationship between theory and observation? Should social scientists have a moral stance toward the individuals and groups that they study? Is this stance appropriate, and would it compromise the researchers' objectivity? These and other questions form part of the theory and practice of social inquiry.

Each of these social inquiry influences is readily seen in the evaluation theories discussed in this section. Influence of the psychology and quantitative sociological disciplines is readily noted in the methods branch. Anthropological thrusts such as "thick description" are evidenced in the work of Stake and Eisner, for example, and attention to concerns raised within the philosophy of science is clearly a part of Scriven's agenda.

METHODS

In the beginning, there was research. And the methods of research dominated the conduct of studies. While most evaluation theorists have methodological concerns and view research as the genesis of program evaluation, one group of theorists has been steadfast in clinging to that orientation. In the social sciences and psychology, this emphasis on research depends on well-designed experimental studies and other controls. Fundamental to these theories are the early work of Donald Campbell (1957) and, in particular, the more popular Campbell and Stanley volume (1966), which defines the conditions for appropriate experimental and quasi-experimental designs.

However, before discussing Campbell and his successors, it is important to address an earlier work that helped to establish a framework for evaluation practice, particularly in education. We have previously noted the broad impact of Ralph Tyler's work in the 1940s on many evaluation theoretical views. We have placed him on the methods branch because we believe that his attention to educational measurement as the essence of evaluation is the most prominent feature of his work.

Ralph Tyler

We view Ralph Tyler's work on *The Eight-Year Study* as one of the major starting points for modern program evaluation (if one can accept the 1940s as modern). As noted in the introductory chapter of this book, Tyler's

work was far-reaching, affecting the work of many future evaluation theorists. Indeed, Madaus and Stufflebeam (1989) credit Tyler for “ideas such as the taxonomic classification of learning outcomes, the need to validate indirect measures against direct indicators of the trait of interest . . . the concept of formative evaluation, content mastery, decision-oriented evaluation, criterion-referenced and objectives-referenced tests” (p. xiii).

Furthermore, Tyler avows that the curricula to be evaluated are based on hypotheses: These are the best judgments of program staff as the best set of procedures for attaining program outcomes. The purpose of evaluation is to validate the program’s hypotheses.

Although he provided a foundation for many theoretical ideas, Tyler’s main focus is on the specification of objectives and measurement of outcomes. He rejects the applicability of norm-referenced tests for program evaluation. He argues that discarding items that were answered correctly by too many, or too few, respondents does not provide the necessary information about what students are learning. Tyler’s point of view has come to be known as *objectives-oriented* (or *objectives-referenced*) evaluation. The approach focuses on (a) formulating a statement of educational objectives, (b) classifying these objectives into major types, (c) defining and refining each of these types of objectives in terms of behavior, (d) identifying situations in which students can be expected to display these types of behavior, (e) selecting and trying promising methods for obtaining evidence regarding each type of objective, (f) selecting on the basis of preliminary trials the more promising appraisal methods for further development and improvement, and (g) devising means for interpreting and using the results (Tyler, 1942, pp. 498-500). Madaus and Stufflebeam (1989) claim that Tyler coined the term “educational evaluation” in the 1930s to describe his procedures—the comparison of (well-stated) intended outcomes (called objectives) with (well-measured) actual outcomes.

OBJECTIVES-ORIENTED THEORISTS

A number of later theoretical works rest heavily on Tyler’s views of evaluation, emphasizing particularly the methodology as objectives-based measurement. Metfessel and Michael’s work (1967) follows Tyler’s evaluation step progression but pays greater heed to expanding the range of alternative

instruments. Hammond (1973) includes Tyler's views as a behavioral objectives dimension that is part of a model that also includes a more precise definition of instruction and the institution. Popham (1973, 1975) follows the Tyler model and focuses primarily on the championing of "behavioral objective specification."

Objectives-oriented evaluation has had a strong continuing influence on education for many decades. As a part of the expansion of Tyler's work, Bloom, Englehart, Furst, Hill, and Krathwohl (1956) developed a taxonomy of educational objectives for the cognitive domain, and Krathwohl, Bloom, and Masia (1964) developed one for the affective domain. These efforts encouraged users to invest a great deal of time and effort into objective specification.

Early work by Popham (1973) called for a narrow scope of educational objectives resulting in a massive number of objectives required to conduct an evaluation and subsequent system overload. Popham, in his later evaluation textbook (1988), recognized this problem and called for a focus on a manageable number of broad-scope objectives and the use of the taxonomies of educational objectives only as "gross heuristics."

Donald Campbell

Shadish et al. (1991) call Donald Campbell the "Methodologist of the Experimenting Society." And indeed he is. Campbell is best known for his path-breaking work on the elimination of bias in the conduct of research in field settings. Most notable from an evaluation perspective are his papers on experimental and quasi-experimental designs for research (Campbell, 1957; Campbell & Stanley, 1966). The focus on experimental designs (and the more practical quasi-experimental designs) is an attempt to "rule out many threats precluding causal inference" (Shadish et al., 1991, p. 122).

There were few papers published during the 20th century that had as great an impact on social science research as Campbell and Stanley's *Experimental and Quasi-Experimental Designs for Research* (1966). Three major areas of social science research design were advanced in this paper. First, the authors explained the conditions necessary to conduct a true experimental study, where randomization is the hallmark. Second, two key notions relevant to this are articulated and explained. They called the degree to which an experiment is properly controlled *internal validity* and referred to how widely applicable the results of an experiment are as *external*

validity. Third, they recognized that experiments are not perfect and that they should not, and cannot, be used in a great many situations. Thus, as an alternative to the true experiment, they describe, in great detail, quasi-experimental designs.

Until their paper was published, in the name of “good science,” the social sciences had used the scientific methods found in the physical sciences. (Experimental designs are still often touted as the most rigorous of all research designs or as the “gold standard” against which all other designs are judged.) Recognized as a striking shift in thinking about social science research, Campbell and Stanley’s paper acknowledges the shortcomings of experiments; in many contexts, experiments are not achievable or desirable. As an alternative to the true experiment, they describe a lengthy set of what they call “quasi-experimental designs.” Quasi-experimental designs were developed to deal with the messy world of field research, where it is not always practical, ethical, or even possible to randomly assign persons to experimental and control groups. Generally speaking, quasi-experimental designs include some type of intervention or treatment and provide a comparison, but they lack the degree of control found in true experiments. Just as randomization is the key to true experiments, lack of randomization is the defining characteristic of quasi-experiments.

The ideas put forth in Campbell and Stanley’s manuscript are now the foundation of almost all social science research methods courses. Perhaps even more remarkable was this paper’s impact on the promotion of new ideas related to research design. It could be argued that the introduction of and favorable response to quasi-experimental designs helped give rise to a climate that was apparently more accepting of alternative thinking about social science research.

Campbell’s work was directed at social science researchers, and it was not until Suchman (1967) saw the relevance of it to evaluation that his name became prominently identified with that field. It is because Campbell’s work on quasi-experimental design precedes Suchman’s application of it to evaluation that we choose to discuss Campbell prior to Suchman.

It should be noted that Campbell (1975a, 1975b) has also written papers indicating the potential appropriateness of qualitative methods as a complement to quantitative experimental methods. Qualitative methods are helpful for evaluation areas such as implementation, identification, and interpretation of side effects. Despite these forays into broadening his methodological

perspectives beyond experimental approaches,¹ Campbell is primarily an experimentalist. His views, along with those of his students, colleagues, and coauthors, have shaped the way that evaluation has been conducted, particularly for those evaluators that we classify on the methods branch of the evaluation tree.

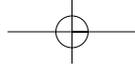
Edward Suchman

An outgrowth of this research tradition, and perhaps the first full-scale description of the application of research methods to evaluation, was a work by sociologist Edward Suchman, who wrote a book titled *Evaluative Research* (1967). The title demonstrates his view of evaluation—as a form of research. Shadish et al. (1991) note that Suchman (1967) is responsible for the evaluation field’s first theoretical integration (p. 32).²

Suchman (1967) distinguishes between evaluation as a commonsense usage, referring to the “social process of making judgments of worth” (p. 7) and evaluative research that uses scientific research methods and techniques. He affirms the appropriate use of the word *evaluative* as an adjective specifying the type of research. He notes, for example, in reviewing early “evaluation guides” put out by various government agencies that their major deficiency is the absence of experimental design. Indeed, Suchman strongly adheres to the importance of conducting evaluative research in scientific ways. To reflect this emphasis, Suchman states that the evaluator must often ask that procedures be altered to secure some form of control or comparative group.

In citing the need for experimental design (and quasi-experimental design), Suchman (1967) highlights the earlier work by Campbell (1957) and the subsequent publication of Campbell and Stanley (1966). Indeed, as Shadish and his colleagues (1991) have pointed out, it was the popularizing of Campbell’s work that helped draw the attention of evaluators to the work and brought Campbell into consideration as an evaluation theorist (and not simply a research methodologist).

The reason Suchman is not defined solely as a researcher applying his work to field situations is his recognition of the administrative context. He comments that the evaluative researcher, in addition to recognizing scientific criteria, must also acknowledge administrative criteria for determining the worthwhileness of doing the study. This had influence on and helped set the tone for many others on the methods branch (such as Rossi, Weiss, Chen, and Cronbach).



But evaluative research is generally applied or administrative research, the primary objective of which is to determine the extent to which a given program or procedure is achieving some desired result. The “success” of an evaluation project will be largely dependent upon its usefulness to the administrator in improving services (Suchman, 1967, p. 21).

It should be noted that Suchman recognizes the role of judgment in evaluation, noting that values and assumptions affect the formulation of goals for evaluative research. However, his extensive discussion of the interlocking of evaluation and administrative roles in program planning and operation places judgment more heavily in the administrative domain—particularly as a second recycling phase following the assessment of outcomes. Thus, while valuing is addressed by Suchman, it plays a subsidiary role to his main emphasis on methods.

Clearly, Suchman’s work had strong influence on the subsequent writings of many who followed, particularly Peter Rossi. There are many examples of his influences, but one in particular is Suchman’s identification of five categories of evaluation: (1) effort (the quantity and quality of activity that takes place), (2) performance (effect criteria that measure the results of effort), (3) adequacy of performance (degree to which performance is adequate to the total amount of need), (4) efficiency (examination of alternative paths or methods in terms of human and monetary costs), and (5) process (how and why a program works or does not work). These multiple emphases also seem evident in Rossi’s work.

Adhering to a strictly chronological sequencing, in this instance, we have placed Suchman on the methods branch following Campbell.

Thomas Cook

The randomized field experiment is seen as the most scientific design available for answering evaluation questions. However, as Cook points out (and as noted previously), this design can be impossible to implement in social and educational program evaluation studies because of contextual factors that are beyond the control of the evaluator. An alternative to this design, Campbell and Stanley’s classic *Experimental and Quasi-Experimental Designs for Research* (1966) delineates quasi-experimental designs that are applicable to evaluation because they address the difficulties of experimental control inherent in field research, specifically random assignment. Generally,



quasi-experiments involve the administration of pre- and posttests and compare groups, but they do not involve randomization.

During the 1970s, Cook and Campbell extended the work of Campbell and Stanley. They contributed new quasi-experimental designs and also called attention to some less recognized threats to internal validity (e.g., resentful demoralization). However, during the mid-1970s, Campbell returned to his initial position supporting the use of the classic experimental design. In fact, he denounced the use of quasi-experimental designs and went so far as to state that perhaps he had committed an injustice by suggesting that quasi-experimental designs were a viable alternative to classic randomized design. Cook, however, remained a proponent of quasi-experiential designs and continued to focus on their use. Today, Cook is generally credited with developing the field of study related to quasi-experimental designs.

Cook expanded on the ideas he put forward with Campbell in several areas. With his interests in methodology, evaluation research, and social reform, Cook was concerned with the contextual factors of the evaluation and the ways that these factors can affect an evaluation research study. Over the years, he wrote extensively on sampling, addressing the issues and alternatives to random selection, methods, the evaluation context, and stakeholder involvement in evaluation studies—all matters that seldom demand attention when implementing the classic experiment. Cook focuses on the importance of using a number of different designs and methods to properly conduct an evaluation. He asserts that it is imperative for evaluators to choose methods that are appropriate to the particular evaluation being conducted and that they take into consideration the context of each evaluation rather than using the same set of methods and designs for all evaluations—a direct attack on the experimental design.

Fairly early on, Cook, unlike other methods-driven theorists of the time, also recognized the importance of involving stakeholders when determining the evaluation questions in order to gather useful information. He explains that a primary problem with evaluations, as he experienced them, is that the evaluators do not consult the decision makers or any other interested parties when forming the research questions that will be addressed by the evaluation. This results in general program goals that are either “operationally vague or of limited importance to many potential users” (Cook & Gruder, 1978, p. 15). Consequently, he suggests that stakeholders should work together with evaluators to decide what an evaluation should examine.

Nonetheless, while Cook expresses concern for identifying users and their interests, he does this within the context of a theorist primarily interested in selecting field-relevant methods that closely resemble the classic experiment. Thus, we place Cook on the methods branch of the tree as someone whose work closely follows Campbell, emphasizing quasi-experimental methods.

Robert Boruch

Robert Boruch's philosophy of evaluation is most similar to the work of Donald Campbell in considering the randomized field experiment as the ideal evaluation approach. Randomized field experiments, however, are distinct from the quasi-experiments put forth by Campbell and Stanley in their seminal piece, *Experimental and Quasi-Experimental Designs for Research* (1966). Boruch (1997) elucidates the distinction between the study designs:

Randomized field tests are also different from *quasi-experiments* [emphasis in original]. The latter research designs have the object of estimating the relative effectiveness of different treatments that have a common aim, just as randomized experiments do, but they depend on methods other than randomization to rule out the competing explanations for the treatment differences that may be uncovered. Quasi-experiments and related observational studies then attempt to approximate the results of a randomized field test. (p. 4)

For Boruch, evaluation is likened to conventional scientific research. It involves the systematic application of rigorous randomized research designs for measuring the extent of a social problem and assessing the implementation, relative efficacy, and cost-effectiveness ratio of social intervention programs (Boruch, Synder, & DeMoya, 2000). Boruch is steadfast in viewing the randomized field experiment as the most effective way of obtaining the least-equivocal estimate of a social program's effects (Boruch et al., 2000).

Boruch, McSweeney, and Soderstrom (1978) explain the randomized experiment as an approach that randomly assigns individuals, or clusters of individuals, to one of two treatment groups. This procedure makes certain that at the outset, in the absence of program-induced differences, groups will be statistically equivalent, thereby maximizing the internal validity of an evaluation. The random assignment approach also avoids the interpretive problems that affect alternative evaluation designs by assuring a fair comparison

between groups (Boruch et al., 2000), ensuring that a legitimate statement can be made about the role of chance in the results.

Boruch (1997) claims that any program, be it social, educational, or welfare, should be studied in a systematic manner, employing randomized controlled experimental methods to gather valid and reliable evidence. It is his commitment to the randomized controlled field experiment³ that defines the Boruch perspective on evaluation. The experimental design approach can “introduce social change in a way that allows one to effectively discern the net effect of the change on important social outcomes” (Berk, Boruch, Chambers, Rossi, & Witte, 1985).

Berk, Boruch, and their colleagues (1985) assert that randomized field experimentation is particularly illuminating when the goal is investigating causal inference and making generalizations to other subjects and settings. They contend that when causal inference is the dominant concern, statisticians widely advocate for experiments that randomly assign subjects to various treatments because random assignment ensures that experimental and control groups are, on the average, equivalent before the treatments were introduced. They maintain that given that social policy experimentation relies upon drawing samples from a known population and then randomly assigning subjects to treatments, sound generalizations will follow (Berk et al., 1985).

We have placed Boruch on the methods branch of the theory tree as a direct continuation of the Campbell experimentation influence. This is in contrast to other disciples of Campbell (such as Thomas Cook) who have advocated more strongly for quasi-experimentation or a number of others influenced by Campbell who have sought to define evaluation more broadly.

Peter Rossi

Peter Rossi is well-known for his highly popular textbook on evaluation. His first edition depicted and relied most heavily on design considerations from Campbell and Stanley (1966), as reflected in Suchman’s influential evaluation book, which focused on evaluation as research. In fact, the title of Rossi’s first edition was *Evaluation: A Systematic Approach* (Rossi, Freeman, & Wright, 1979).

Now in its sixth edition (with modified coauthors), it is still a very popular and influential text. Over time, however, because of the influences of other evaluation theorists, the text now includes discussions of qualitative data

collection, evaluation utilization, the role of stakeholders, and so on. It is interesting to note the ways in which other perspectives and the changing nature of the field modify views.

Rossi's earlier writings stressed the use of the experimental design. However, to describe him as exclusively an experimentalist would be erroneous. He contributed significantly to the conceptualizations of theory-driven evaluation and of comprehensive evaluation, which he describes as "the systematic application of social research procedures in assessing conceptualization and design, implementation, and utility of social intervention programs" (Rossi & Freeman, 1985, p. 19). However, as is evident from this quote, while he has moved away from considering the experimental model as the sine qua non of evaluation, he still clearly views evaluation as social research.

Rossi's departure from the randomized experiment was initiated in part by his broadening vision of evaluation. Concerning himself with issues such as internal and external validity, implementation, and use, his ideas evolved, eventually, to a place that some suggest was so comprehensive that the approach he suggested was virtually impossible to implement. Rossi's response to this criticism led him to develop "tailored evaluations" that assisted in refining the evaluation focus. In this work, Rossi suggests that methods be implemented that are tailored to the stage of the program, that is, "fitting evaluations to programs" (Rossi & Freeman, 1985, p. 102). For example, if commissioned to conduct an evaluation of a program that is well under way, the evaluation would be designed to focus on current work rather than on the development of the program. That is not to say that program development would not be addressed by the evaluation; rather, it would receive less attention than if the commencement of the evaluation was concurrent with the start of the program.

Rossi, along with Weiss and Chen, provided some of the foundational thinking about theory-driven evaluation. This approach, which is discussed in depth later in this chapter in relationship to Chen, can be seen as a particular type of comprehensive evaluation. Briefly, theory-driven evaluation involves the construction of a detailed program theory that is then used to guide the evaluation. Rossi maintains that this approach helps to reconcile the two main types of validity, internal and external. Since, for a variety of reasons, it is difficult to design studies that emphasize both, many theorists believe that one must be traded for the other. Nevertheless, both are important. Rossi's earlier work focusing on the use of the experimental design was obviously concerned more with internal than external validity. However, he began to question whether

internal validity should have priority. Rossi asserted that theory-driven evaluation is to a certain extent a fusion of Campbell's and Cronbach's approaches.⁴ This fusion allows the evaluator to address internal validity through building a specified program model and implicit external validity by investigating the ways in which the program is similar to and different from the future program to which the evaluation is to be generalized.

Shadish et al. (1991) say that Rossi presents a multifaceted, broad-based point of view to appeal to a wide audience. In our view, however, Rossi has demonstrated a dominant predilection for the purity of experimental and quasi-experimental methods. Thus, we believe it more accurate to say that the recent edition of Rossi's text (Rossi, Freeman, & Lipsey, 1999) reflects a concern for demonstrating great breadth and for the depiction of evaluation possibilities to capture the interests of broader audiences.

Huey-Tsyh Chen

Huey Chen is most influential in developing the concept and practice of theory-driven evaluation. He has frequently coauthored articles with Peter Rossi and, in addition, has authored a book and articles on the topic. Chen acknowledges the attractiveness of controlled experiments in estimating net effects through randomization. But when no effect is shown in a controlled experiment, Chen points out that there is no indication as to whether failure is due to, for example, poorly constructed causal linkages, insufficient levels of treatment, or poor implementation. Chen proposes a solution to this dilemma:

We have argued for a paradigm that accepts experiments and quasi-experiments as dominant research designs, but that emphasizes that these devices should be used in conjunction with a priori knowledge and theory to build models of the treatment process and implementation system to produce evaluations that are more efficient and that yield more information about how to achieve desired effects. (Chen & Rossi, 1983, p. 300)

Chen recognizes the dominance of the experimental paradigm but strongly believes that it must be supplemented by the development of theoretical models of social interventions: "An unfortunate consequence of this lack of attention to theory is that the outcomes of evaluation research often provide narrow and sometimes distorted understandings of programs" (Chen & Rossi, 1983, p. 284).

Chen is concerned with identifying secondary effects and unintended consequences. This is similar to Scriven.⁵ Chen, however, relies on social science theory for the identification of potential areas for investigation. In pursuing theory-driven evaluation, Chen acknowledges the potential effects of a program as related to its official goals, but he also sees the necessity for using social science knowledge and theory related to the subject matter in question to identify effects that go beyond those goals. The theories that he seeks to construct are not global or grand, but “plausible and defensible models of how programs can be expected to work” (Chen & Rossi, 1983, p. 285).

Chen acknowledges the validity questions raised in following Campbell’s and Cronbach’s approaches but believes that they can be resolved. “We are not convinced that the trade-off problem between internal and external validity, which is so sharply portrayed by Campbell, Cronbach, or others, that dealing with one type of validity must seriously sacrifice the other types of validity” (Chen & Rossi, 1987, p. 97). He believes that theory-driven evaluation enhances the possibility of achieving each type of validity. This is accomplished by doing randomized experiments and simultaneously studying causation within the context of constructed theories (usually using structural equation modeling).

In conducting evaluations, Chen places his focus not only on the use of social science knowledge but also on traditional research methods. Chen’s place on the evaluation tree is on the methods branch, his work rooted in the methodological traditions of social science.

Carol Weiss

The title of Carol Weiss’s chapter in *The National Society for the Study of Education (NSSE) Yearbook on Evaluation* (Weiss, 1991) is indicative of her appropriate placement on this evaluation theory tree: “Evaluation Research in the Political Context: Sixteen Years and Four Administrations Later.” “Evaluation research” are the code words for emphasis on what we call “methods.” The same or similar terms are used in titles by Suchman and Rossi and in many of Weiss’s other publications (see Weiss, 1972a, 1972b, 1973).

Weiss’s early evaluation work was influenced by research methodologists, political theorists, and expositors of democratic thought (e.g., Rousseau, the *Federalist Papers*). Weiss’s firm commitment to evaluation as being synonymous with carefully conducted research is documented in many of her

writings. In fact, her first evaluation text, *Evaluation Research*, was intended to be a supplementary text for research methods courses. Shadish et al. (1991) point out that Weiss's early writings focused almost exclusively on "traditional experimental methods in evaluation" (p. 183). While in later writings, she moved somewhat away from this as her main emphasis, she still clung to a strong insistence on methodologically sound research. "Evaluation is a kind of policy study, and the boundaries are very blurred. . . . I think we have a responsibility to do very sound, thorough systematic inquiries" (Weiss, cited in Alkin, 1990, p. 90).

The reference to "political context" in Weiss's chapter title demonstrates her added perspective that recognizes evaluation as a political activity. It is political in that the activities of evaluation, both as a process and in terms of research findings, affect and are affected by the political context. Weiss (1991, p. 214) is particularly pointed in recognizing that the process of conducting an evaluation affects a political situation in which there are vested interests, negotiation, constituents, critics, and so on.

Politics intrudes on program evaluation in three ways: (1) programs are created and maintained by political forces; (2) higher echelons of government, which make decisions about programs, are embedded in politics; and (3) the very act of evaluation has political connotations. (Weiss, 1991, p. 213)

Weiss recognizes the complexity of political context and was among the first to note that evaluations do not take place in a political vacuum. Furthermore, she has been adamant in noting that evaluation results are not necessarily (or generally) heeded in isolation of other input or at the time of evaluation reporting. She is known for, among many other contributions, introducing the term *decision accretion* into the evaluation lexicon. By this, she means that decisions are the result of "the build-up of small choices, the closing of small options and the gradual narrowing of available alternatives" (Weiss, 1976, p. 226).

Given the belief that decisions accrete, Weiss has studied policy research to determine the characteristics of evaluative research work that have longevity—that is, those that would be considered credible during an accretion process. She concludes that research quality is of greatest importance: "Research quality is of value not only because of adherents to the norms of science, but perhaps more importantly because it increases the power of research as ammunition in intra-organizational argument" (Weiss & Bucuvalas,

1980, p. 256). It is this continued concern for “careful and sound research” (Weiss, 1981, p. 400) that leads to her placement on the methods branch of the tree. Since she places limits on the likelihood of instrumental use and instead views evaluation’s greatest contribution as conceptual or enlightenment use, she does not fit within the use branch, which is instrumental in its orientation.

As Shadish et al. (1991) note about Weiss, “When forced to choose between traditional scientific rigor versus other goals such as use, she seems to lean in the former direction” (p. 208).

Lee J. Cronbach

Lee J. Cronbach is one of the methodological giants of our field. His methodological contributions include Cronbach’s coefficient alpha, generalizability theory, and notions about construct validity. Thus, he has strong evaluation roots in methodology and social science research that led us to place him on the methods branch of the theory tree. Yet there are elements in his work that reflect a broader view than the Campbell-like social science research mode. To begin with, Cronbach was a student of Ralph Tyler’s, lending his work a kind of field orientation not present in the work of many social science researchers. Furthermore, his association with more policy-research-oriented Stanford University colleagues, notably in his book *Toward Reform of Program Evaluation* (Cronbach & Associates, 1980), helped to establish his concern for evaluation’s use in decision making.

Following on the work of Weiss, Cronbach rejects the simplistic model that assumes a single decision maker and “go/no-go” decisions. Instead, he views evaluation as an integral part of policy research focused on policy-shaping communities and necessitating potential political accommodation. Thus, he is more oriented toward evaluation utilization for enlightenment purposes than toward a concern for more immediate instrumental use. He does, however, affirm the evaluator’s active role in providing data to local decision makers for instrumental use (in accordance with contractual obligations).

To define the manner in which evaluators might most productively enhance enlightenment use, Cronbach coins a set of symbols to define the domains of evaluation (Cronbach, 1982). These domains consist of *units* (populations), *treatments*, *observations* (outcomes), and *settings*. The sample of this domain examined by the evaluator is referred to by the acronym “utos,” and it represents the larger population from which sampling took place,

which is referred to as “UTOS.” Cronbach maintains that the most important area of concern for evaluation is external validity, which he refers to as the plausibility of the conclusions to UTOS that are manifestly different from the population under study. He refers to this manifestly different population as “*UTOS” (pronounced “star-UTOS”). In the concern for evaluations contributing to enlightened discussion, Cronbach focuses on what he refers to as “bandwidth,” by which he means that it is more important for the evaluation to focus on a broad range of relevant issues than to achieve absolute fidelity (accuracy) on a small number of issues. Cronbach’s concern about generalizing to *UTOS leads him to reject Campbell and Stanley’s emphasis on experimental design and Scriven’s focus on comparison programs. He proposes that generalization to *UTOS can be attained by extrapolating through causal explanation, either using causal modeling or the “thick description” of qualitative methods. Furthermore, it is sometimes beneficial to examine subpopulations (sub-UTOS). Thus, focusing on the subset of data for a particular group might enable generalization to other domains. Furthermore, he seeks to capitalize on naturally occurring variability within the sample as well as consequences of different degrees of exposure to treatments. This work is an extension of earlier research on aptitude-treatment interactions conducted jointly with Snow (Cronbach & Snow, 1977).

Cronbach displays sensitivity to the values of the policy-shaping community and seeks to incorporate their views on the evaluation questions most worth asking. This is done systematically with an eye to what will contribute most to generalization: issues receiving attention from the policy-shaping community; issues relevant in swaying important (or uncommitted) groups; issues having the greatest uncertainty; and issues that would best clarify why a program works. Shadish et al. (1991) make the following keen distinction between Cronbach and several other major theorists:

[Cronbach views] evaluators [as] educators rather than [as] the philosopher-kings of Scriven, the guardians of truth of Campbell or the servants of management of Wholey. (p. 340)

As such, Cronbach does not aspire to the instrumental use of evaluation as do Wholey and others on the decision-making/use branch. And, as we have pointed out, he does not adhere to the strict experimental controls that Campbell advocates. Likewise, he does not call upon the evaluator to impose

his or her own value system on the program or to reach summary judgments about programs. Rather, he sees the evaluator's role as providing "readers" with information that they may take into account when forming their own judgments.

We have placed Cronbach on the methods branch of the evaluation tree, but branching off in the direction of the use branch. This placement in part reflects the influence of Weiss on his ideas with respect to enlightenment use directed toward policy-shaping communities.

VALUING

Out of the trunk of social inquiry has grown a branch of evaluators who focus on concerns related to valuing or the making of judgments. Theorists on this branch believe that what distinguishes evaluators from other researchers is that evaluators must place value on their findings and, in some cases, determine which outcomes to examine. This branch of evaluation has as its primary mainstay Michael Scriven, whose early work (1967, 1972a) was a brilliant exposition that largely defined the field in terms of its conceptual underpinnings and procedures.

Michael Scriven

Scriven's major contribution is the way in which he adamantly defined the role of the evaluator in making value judgments. Shadish et al. (1991) noted that Scriven was "the first and only major evaluation theorist to have an explicit and general theory of valuing" (p. 94). Scriven is unequivocal in his position that society requires valuing and it is the role of the evaluator to do that job. He maintains that there is a science of valuing and *that* is evaluation. Scriven (1986) notes, "Bad is bad and good is good and it is the job of evaluators to decide which is which" (p. 19). Scriven (1983) notes that the greatest failure of the evaluator is in simply providing information to decision makers and "passing the buck [for final judgment] to the non-professional" (p. 248).

The evaluator, in valuing, must fulfill his or her role in serving the "public interest" (Scriven, 1976, p. 220). By public interest, Scriven does not restrict the evaluator's responsibility simply to clients, users, or stakeholders, but to all potential consumers. Indeed, he views the evaluator's role in valuing as

similar to producing a report for *Consumer Reports* in which the evaluator determines the appropriate criteria for which judgments are to be made and then presents these judgments for all to see. As in *Consumer Reports*, there is the necessity for identifying “critical competitors,” that is, competing alternatives. Comparisons are key in making value judgments, and the evaluator has the responsibility for identifying the appropriate alternatives. Just as *Consumer Reports* would identify “midsize cars” to be evaluated, so, too, would the evaluator seek out similar entities for evaluation.

Scriven adamantly states that it is not necessary to explain why a program or product works to determine its value. Nevertheless, he introduces an alternative to experimental and quasi-experimental design called the “modus operandi” (MO) method (Scriven, 1991, p. 234). Analogous to procedures used to profile criminal behavior,

The MO of a particular cause is an associated configuration of events, processes, or properties, usually in time sequence, which can often be described as the *characteristic causal chain* (or certain distinctive features of this chain) connecting the cause with the effect. (Scriven, 1974, p. 71)

In a manner similar to that which Scriven uses to grade critical competitors, the MO method requires that the evaluator first develop an exhaustive list of potential causes and then narrow down the potential choices. The list of potential causes is narrowed down in two steps. In the first step, the evaluator determines which potential causes were present prior to the effect. In the second step, the evaluator can determine which complete MO fits the chain of events and thus determine the true cause. To ensure accuracy and bias control, the evaluator looks for instances of “co-causation and over determination,” and calls in a “goal-free or social process expert consultant to seek undesirable effects” (Scriven, 1974, p. 76). Scriven believes that ultimately, the evaluator is able to deliver a picture of the causal connections and effects that eliminate causal competitors without introducing evaluator bias.

Furthermore, Scriven (1972b) advocates “goal-free evaluation” in which the evaluator assumes the responsibility for determining which program outcomes to examine, rejecting the objectives of the program as a starting point. He maintains that by doing so, the evaluator is better able to identify the real accomplishments (and nonaccomplishments) of the program. An essential element of Scriven’s valuing is the determination of a single value judgment of the program’s worth (“good” or “bad”). In requiring a synthesis of

multiple-outcome judgments into a single value statement, Scriven is alone among evaluation theorists.

In both goal-free evaluation and in the synthesis stage, Scriven justifies his point of view by relying on the extent to which the program is able to meet “needs.” Needs are the presumed cost to society and the individual and are determined through a needs assessment. Shadish et al. (1991) maintain that Scriven’s conception of valuing is dependent on his definition of needs and note that his “conception of needs implies a prescriptive theory of valuing and that he disparages descriptive statements about what people think about the program” (p. 95). Moreover, they maintain that his needs assessment is not independent of the views of the evaluator and that failing to directly reflect the views of stakeholders inhibits the potential use of evaluation findings in policymaking. Scriven is apparently unconcerned by this, maintaining an idealist position of determining “truth” whose revelation is sufficient. Scriven’s unique training in philosophy, mathematics, and mathematical logic provides him with the assurance that he can make sound, unbiased judgments.

Scriven is the leading theorist of the valuing perspective and provides the “spiritual guidance” and direction for others depicted on this branch.

Elliot Eisner

Elliot Eisner’s approach to evaluation is very different from Scriven’s, yet they share similarities with respect to the importance they give to valuing and judging. Indeed, Eisner (1998) notes that evaluation “concerns the making of value judgments about the quality of some object, situation or process” (p. 80). This is why Eisner, too, is on the values branch.

Eisner first presented his views on what he called “educational connoisseurship” in an issue of the *Journal of Aesthetic Education* (1976) and has subsequently expanded on those views (1985, 1991a, 1991b, 1998). His evaluative viewpoint is a direct response to negative views on traditional approaches focusing on educational outcomes, whether measured by standardized tests using principles of psychological testing or by criterion-referenced testing procedures, which were a major part of objective-based evaluation (see Tyler, 1942, and others). Eisner’s rejection of “technological scientism” includes a rejection of the extensive use of research models employing experimental and quasi-experimental design, which are heavily (if not exclusively) dependent on quantitative methods. Eisner notes that things that matter cannot be

measured quantitatively (1976). He argues that while quantitative techniques can provide some useful information, "Evaluation requires a sophisticated, interpretive map not only to separate what is trivial from what is significant, but also to understand the meaning of what is known" (Eisner, 1994, p. 193).

Drawing heavily from his prior experience as a curriculum expert and as an artist, Eisner uses the role of critics in the arts as an analogy for an alternative conception of evaluation. Central to his evaluation theoretic views are the twin notions of connoisseurship and criticism. To be a connoisseur is to have knowledge about what one sees, to have the ability to differentiate subtleties and to be aware of and understand the experience. Eisner (1991b) notes that "a connoisseur is someone who has worked at the business of learning how to see, to hear, to read the image or text and who, as a result, can experience more of the work's qualities than most of us" (p. 174).

Criticism is making the experience public through some form of representation. Eisner describes three aspects of criticism. First, there is critical description, in which the evaluator draws upon her senses to describe events, reactions, interactions, and everything else that she sees. By doing this, the evaluator portrays a picture of the program situation, frequently imagining herself as a participant and drawing upon her senses to describe the feeling in the participant's terms. The second dimension of criticism is expectation, in which the evaluator tries to understand or make sense of what she has seen. She asks, "What is the meaning of this?" "What message are participants getting?" In responding to these questions, the evaluator must draw upon her own experience and place herself in the minds of participants. It is in this role that Eisner emphasizes the importance of necessary background knowledge, since a good critic cannot interpret or evaluate what she has not experienced herself. The critic must ask, "What is the value of what is happening?" Eisner (1991b) notes that valuing is a critical element in the process and further elaborates,

The essence of perception is its selectivity; the connoisseur is as unlikely to describe everything in sight as a gourmet chef is to use everything in his pantry. The selective process is influenced by the value one brings to the classroom. What the observer cares about, she is likely to look for. . . . Making value judgements about the educational import of what has been seen and rendered is one of the critical features of educational criticism. (p. 176)

Like Scriven, Eisner considers the valuing role important and expands the role not only to include making final judgments about data (or observations) but also to include making judgments about what questions to ask and what to focus on. Eisner recognizes that goals, possible outcomes, and observations are limitless. Eisner and Scriven each place the determination of what is important to look at in the hands of the evaluator. Note that while Scriven's evaluator draws his or her expertise from philosophy, logic, mathematics, and scientific methods, Eisner's evaluator draws from connoisseurship and experience or "the enlightened eye" (1991a). Scriven's methodology draws heavily from the quantitative research traditions, while Eisner's, as we have seen, relies almost exclusively on qualitative methods, yet each places importance on the evaluator's role in valuing and judging.

Thomas Owens/Robert Wolf (Adversary Evaluation)

A theoretic approach to evaluation that clearly reflects a valuing orientation is generally referred to as "adversary evaluation." While we are not aware of any current advocates of this point of view, nonetheless it is worth noting since it is unique and it attracted a good deal of attention in the 1980s. Based on a concern about the unintentional bias of evaluators in their choice of outcomes to examine or in their choice of the basis for valuing, Tom Owens and Robert Wolf each, separately, developed evaluation approaches that are typically classified as *adversary evaluation models*. The method they suggest to reduce potential bias is to employ evaluators to represent two opposing views (advocates and adversaries). The two sides then agree on issues to be addressed. A common database is collected. The adversary procedures primarily relate to the way in which the evaluation data are synthesized and recorded. The two sides must agree beforehand on the issues agreed upon and those under contention. Each side aggregates and synthesizes the common data as well as additional data—primarily interviews—that best present their argument.

Owens (1973) describes what he calls a "modified judicial model" as a basis for his adversary approach. Worthen and Sanders (1987) note that that includes "pre-trial conferences, cases presented by the 'defense' and 'prosecution,' a hearings officer, a 'jury' panel of educators, charges and rebuttals, direct questioning and redirected questions, and summaries by the prosecution and defense" (p. 115).

Wolf proposes a judicial model with elements similar to a “trial by jury” as a basis for his approach (1975). He defines a variety of key roles to be played, roughly corresponding to those found in a courtroom setting, and he uses a system of evidentiary rules and procedures as guides for fact-finding and inquiry. Wolf (1979) maintains that judgments could be rendered that are fair, honest, and best serve the public interest. Wolf prefers to identify his paradigm for educational evaluation as the “judicial evaluation model” (JEM) and criticizes those theorists who refer to it as the “adversarial model.” Nonetheless, because of the similarity of Wolf’s approach to Owens’s, we list these two theorists together.

The adversary model was used as the basis for a major evaluation of a large-scale team-teaching program in K-3 classes in Hawaii. It was described in a report of the Northwest Regional Educational Laboratory (1977) and was critiqued by two of the adversarial participants (Popham & Carlson, 1977).

We classify Owens’s and Wolf’s approaches on the valuing branch of the evaluation family tree because their work focuses primarily on the manner in which judgment would be reached in an evaluation. Unlike the approaches of Scriven or Eisner, where the valuing is done by a single evaluator (or evaluation team), in adversary-oriented evaluation approaches, there are two distinct sets of evaluators placing value on data, and these sets are then thrust into adversarial positions. The ultimate judge has two value interpretations to choose from, perhaps accepting a bit of each.

Robert Stake

We found Robert Stake the most difficult of the theorists to categorize. His views on responsive evaluation have had substantial impact on the evaluation theorists on the use branch as well as on those on the valuing branch. He has been a strong advocate of case studies and of the evaluator producing a “thick description.” A serious argument could be made for placing him on any of the three branches. As House points out (2001), the essential components of Stake’s responsive evaluation are (a) the belief that there is no true value to anything (knowledge is context bound), (b) the belief that stakeholder perspectives are integral elements in evaluations, and (c) the belief that case studies are the best method for representing the beliefs and values of stakeholders and of reporting evaluation results. Actually, Stake’s views are broader than that. He maintains that case study is necessary for context and

activity description but might, for some audiences, be a poor way to judge quality. Stake's (2001) own words often portray the image of a theorist who, in our schema, should be placed on the valuing branch: "Now I think of the judgmental act as part of the descriptive, part of the observational act" (p. 10).

In other words, Stake maintains that seeing and judging the evaluand regularly are part of the same act and that the task of evaluation is as much a matter of refining early perceptions of quality as of building a body of evidence to determine level of quality.

Stake maintains that there are multiple realities and that stakeholder perspectives need to be represented within the evaluation, but he just as firmly believes that stakeholders do not participate in the evaluation in the same way that participant theorists would have them do. He is opposed to stakeholder participation and asserts that evaluation is the job of the evaluator (Alkin, Hofstetter, & Ai, 1998, p. 98). As House (2001) notes in discussing Stake, "What goes into the report and what the evaluator responds to is at the discretion of the evaluator" (p. 26). Stake's view on how evaluators draw conclusions is that "ultimately the evaluator decides, which must mean that judgment is dependent on the evaluator's values and personality" (House, 2001, p. 28). However, judgment is not dependent on those alone. Stake (2000) notes that it is the evaluator's job "to hear the [participants'] pleas, to deliberate, sometimes to negotiate, but regularly non-democratically, to decide what their [the participants'] interests are" (p. 104). From these statements and views, one could make a strong case for including Stake on the judgment branch.

On the other hand, Stake (1975) has noted that "whatever consensus in values there is [among participants] . . . should be discovered. The evaluator should not create a consensus that does not exist" (pp. 25-26). Furthermore, Stake notes, "The reader, the client, the people outside need to be in a position to make their own judgments using grounds they have already, plus the new data" (Tineke & Stake, 2001, p. 10).

Stake's attention to the notion of stakeholder participation (albeit somewhat limited compared with others' notions of it) and the visibility that he accords these ideas have had a strong influence on theorists on the use branch of this tree. For that reason, we considered placing him on the use branch. While the primary orientation of Stufflebeam and others on the use branch is evaluation for decision making, focusing typically on program administrators as the decision makers, Stake extends the user group to a variety of stakeholders. Although he gives attention to how audiences may understand the

reporting, he says very little about how the evaluation findings might be used. One can see the easy transition from the stakeholder concept to Patton's "primary intended users" (a comparatively minor shift from Stufflebeam's position). On the other hand, in extending further out on the branch, the stakeholder concept, usually modified to include stakeholder participation as it is written about by Stake, has had a major impact on other use theorists.

Alternatively, if one were to place Stake on the valuing branch, there is an easy transition to Ernest House, who embraces the stakeholder notion but adds the perspective of giving voice to the underrepresented whose voices are not usually heard. Moreover, the notions of multiple realities and the role of the evaluator in providing the basis for alternative values to be expressed rings deeply in the work of Egon Guba and Yvonna Lincoln (although, unlike Stake, they view their role as informing judgments). Like Cronbach, Stake acknowledges the teaching role of evaluators as important, not only for delivering judgments but also by providing learning opportunities for those who will judge.

In designating Stake's place on the evaluation tree, we were persuaded by Stake's own statements above as well as his statement, "I never give up the evaluator's responsibility to provide summary judgment, but I provide them softly framed so as to encourage the reader's own interpretations" (Stake, 2001, p. 10).

Thus, we have placed Stake on the valuing branch but acknowledge his substantial contribution to theorists of different types.

Barry MacDonald

Barry MacDonald has been placed on the valuing branch. He, like Lincoln and Guba, recognizes the multiplicity of perspectives held by stakeholders and believes the evaluator's duty is to present the values of differing stakeholders. MacDonald (1979) depicts the evaluator as a "negotiator of perspectives" (p. 131).

[The evaluator] will refrain from making judgments of merit, but will instead seek to make a program accessible to those who wish to judge its assumptions, its operations, or its achievements. . . . This does not mean that the evaluator will retain the right to make judgment. Rather he will collect and communicate the alternative definitions, perspectives and judgment held by people in and around the program. (pp. 127-128)

MacDonald argues for awareness of the differing political contexts and offers a classification system for each. He describes evaluations as occurring in bureaucratic, autocratic, or democratic contexts. Bureaucratic evaluations are done for the sake of the people in power. In this context, the evaluator accepts the values of those who hold office, and offers information that will help them to accomplish their policy objective; "Its key justificatory concept is 'the reality of power'" (MacDonald, 1977, p. 224).

Autocratic evaluations are similar to bureaucratic evaluations with the key exception being the independence of the evaluator. In this context, the evaluator "focuses on issues of merit and acts as expert advisor. His technique of study must yield scientific proof. . . . Its key justification concept is 'the responsibility of office'" (MacDonald, 1977, p. 224).

Democratic evaluations are conducted for the sake of informing the community. Here, the evaluator "recognizes the value pluralism and seeks to represent a range of interests in its issue formulation. His main activity is the collection of definitions and reaction to the program. . . . The key justificatory concept is 'the right to know'" (MacDonald, 1977, p. 224).

MacDonald (1979) advocates the use of democratic evaluations due to their ability to "portray the multiple reality of a program with justice and truth" (p. 131), leaving decision makers with a rich array of perspectives and judgments to consider when making a decision. He maintains that conducting democratic evaluations allows the evaluator to "adopt a [political] stance that is not vulnerable to criticism on the grounds of partiality" (MacDonald, 1977, p. 108). In this role, evaluators can facilitate negotiations between differing standpoints, perspectives, and values.

Thus, we have placed Barry MacDonald on the valuing branch of the tree due to his concern for representation of the values of others during the evaluation process.

Ernest House

Ernest House recognizes that evaluation serves the purpose of providing information to decision makers so that they can determine the legitimate allocation of vital resources. He denounces as inadequate the utilitarian framework that underlies key theories of evaluation (primarily those of writers we classify on the methods branch and early theories of those we

classify on the use branch). House (1991) notes that “utilitarianism is a moral theory which holds that policies are morally right when they promote the greatest sum total of good or happiness from among the alternatives” (p. 235). In pursuit of social justice, he recommends attempts at being responsive to stakeholders as being superior to prior evaluation conceptions. He also deplores the lack of value neutrality in stakeholder approaches, which he says results from the general lack of full inclusion of the represented interests of the poor and powerless in stakeholder groups (pp. 239-240).

House comes to these views by drawing on Rawls’s justice theory (Rawls, 1971). House (1991, 1993) argues that evaluation is never value neutral; it should tilt in the direction of social justice by specifically addressing the needs and interests of the powerless. House’s evaluator is thus faced with the task of understanding the needs and positions of various stakeholder groups, especially the poor and powerless, and of balancing this information with his or her perception of social justice. In doing this balancing, the evaluator would shape the kind of information collected, as well as its analysis. In essence, evaluators cast themselves in the role of spokespeople or representatives of the poor and powerless.

It is important to note that for House, the role of evaluator is not to define value in terms of good or bad, as Scriven does, but in terms of right, fair, and just. In this sense, the value judgments accorded to the least advantaged would receive the utmost importance. Thus, both Scriven and House place the valuing component in a position of eminence, but they do so with a substantially different emphasis.

It is informative to examine what House refers to as “ethical fallacies” within evaluation. They are *clientism* (taking the client’s interest as the ultimate consideration); *contractualism* (adhering inflexibly to the contract); *managerialism* (placing the interest of the managers above all else); *methodologicalism* (believing that proper methodology solves all ethical problems); *pluralism/elitism* (including only the powerful stakeholders’ interests in the evaluation); and *relativism* (taking all viewpoints as having equal merit.) By referring to these as ethical fallacies, House disassociates himself with evaluation theoretic goal views that are consistent with each. For example, the first three of these can easily be thought of as associated with the use branch of the theory tree (particularly the early part of the branch). Methodologicalism is clearly associated with the methods branch, and pluralism/elitism and relativism are associated with various stakeholder approaches. House’s emphatic

concern with values and judgment in his theoretical views firmly establishes his place on the valuing branch of the evaluation tree.

Egon Guba and Yvonna Lincoln

We have placed Guba and Lincoln (*Fourth Generation Evaluation*) on the valuing branch. While, as with many theorists, they share procedures common with others on a different branch (in their case, the use branch), their orientation is focused on valuing. Unlike others on this branch, however, such as Scriven or Eisner, who view the evaluator as the “valuer,” Guba and Lincoln view stakeholders as the primary individuals involved in placing value. This viewpoint rests on the belief that instead of there being one reality, there are multiple realities based on the perceptions and interpretations of individuals involved in the program to be evaluated. Thus, Guba and Lincoln believe that the role of the evaluator is to facilitate negotiations between individuals reflecting these multiple realities.

Fourth Generation Evaluation (Guba & Lincoln, 1989) is based upon a constructivist paradigm: That is, in place of the existence of a single reality, individuals “construct” their perceptions of reality. The role of the constructivist investigator is to tease out these constructions and “to bring them into conjunction . . . with one another and with whatever information . . . can be brought to bear on the issues involved” (p. 142). More specifically, evaluators are “orchestrators of a negotiation process that aims to culminate in consensus on better informed and more sophisticated constructions” (p. 110).

Guba and Lincoln (1989) claim that *Fourth Generation Evaluation* is “a marriage of responsive focusing—using the claims, concerns and issues of stakeholders as the organizing elements—and constructivist methodology aiming to develop judgmental consensus among stakeholders who earlier held different, perhaps conflicting, emic constructions” (p. 184). Thus, they use maximum variation sampling to identify the broadest scope of stakeholders who are interviewed sequentially in order to place on the table the great variety of individual constructions. These are part of what they refer to as “the hermeneutic circle,” which defines a continuous interplay of data collection and analysis. Ideally, out of this process, a joint construction will begin to emerge. Additional roles for the evaluator include “testing and enlarging within-group constructions by introducing new or additional information”; “sorting out resolved CC & I [claims, concerns and issues]”; “prioritizing

unresolved CC & I"; "collecting information bearing on unresolved CC & I"; "preparing an agenda for negotiation"; "carrying out a negotiation"; and "reporting via the case study—the joint construction as product" (Guba & Lincoln, 1989, p. 185).

Aside from the emphasis on bringing together constructions to facilitate valuing by stakeholders, there are numerous places in the process where the evaluator himself or herself engages in valuing. Indeed, according to these theorists, "'facts' and 'values' are interdependent. 'Facts' have no meaning except within some value framework; they are value-laden. There can be no separate observational and valuational languages" (Guba & Lincoln, 1989, p. 105).

A number of Guba and Lincoln's (1989) comments seem to validate the point that fourth generation evaluators do play at least a moderate personal role in valuing. Note the following:

As the several constructions begin to take shape, however, certain elements will seem to be more salient than others (and will probably first appear this way to the inquirer). (p. 153)

The human is the instrument of choice for the constructivist for only a human "can enter a context without prior programming, but . . . after a short period begin to discern what *is* salient (in the emic view of the respondents). (p. 175)

The inquirer's own etic (outsider) construction may be introduced for critique . . . so long as all respondents have the opportunity to criticize the inquirer's formulations as they must do with their own. (p. 154)

The constructivist moves into a situation without prior propositional formulations in mind. . . . How can he or she go about sensing out what to examine? The constructivist's answer to that question is to bring one's tacit knowledge to bear. . . . It is precisely this same tacit understanding of a situation that serves the constructivist in the beginning stages of an inquiry. (pp. 176-177)

Positivists ask whether the evaluator does not abrogate objectivity when he or she becomes involved in shaping others' constructions. (p. 207)

Thus, it is clear that Lincoln and Guba's evaluation theoretical views fit nicely on the valuing branch of the evaluation theory tree. It is particularly interesting to observe among the various theorists on this branch the great diversity in the manner in which valuing takes place, yet all share a primary

concern for valuing either by the evaluator or by the evaluator in a role designed to facilitate the valuing by others.

USE

The use branch began its growth with what are often referred to as “decision-oriented theories.” Decision-oriented theorists felt it was critical to conduct evaluations that were designed specifically to assist key program stakeholders in program decision making. Such stakeholders are most often those that commission the evaluation. Stufflebeam’s CIPP model is one of the most well-known of these theories. Based on empirical knowledge, utilization theorists built on the notions put forth in decision-oriented theories. This class of theories is concerned with designing evaluations that are intended to inform decision making, but it is not their only function to ensure that evaluation results have a direct impact on program decision making and organizational change.

Daniel Stufflebeam

Daniel Stufflebeam, along with Guba, initially developed the CIPP model as an approach to evaluation focused on the decision-making process. CIPP is an acronym for four types of evaluation: *context*, *input*, *process*, and *product*. Context evaluation involves identifying needs to decide upon program objectives. Input evaluation leads to decisions on strategies and designs. Process evaluation consists of identifying shortcomings in a current program to refine implementation. Product evaluation measures outcomes for decisions regarding the continuation or refocus of the program.

Stufflebeam describes the CIPP model of evaluation as a cyclical process. The key strategy is to work with a carefully designed evaluation while maintaining flexibility. According to Stufflebeam (1983), evaluators must view design as a process, not a product. Evaluations should provide a continual information stream to decision makers to ensure that programs continually improve their services. To improve services, evaluations should aid decision makers in allocating resources to programs that best serve clients.

Stufflebeam (2000) often keys his evaluations to *The Program Evaluation Standards* (Joint Committee for Educational Evaluation, 1994). Focusing on

four domains related to practice, *The Program Evaluation Standards* describes the professional standards by which evaluators should conduct their work. The four domains are *utility*, *feasibility*, *propriety*, and *accuracy*. The utility standards are intended to ensure that an evaluation will serve the information needs of intended users. The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal. The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due respect for the welfare of those involved in the evaluation, as well as of those affected by its results. The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated. Stufflebeam also believes that evaluations should be grounded in clear contracts. Citing the feasibility standard of formal contracts, he advises evaluators and clients to reach formal written agreements that detail “what is to be done, how, by whom, and when so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it” (Stufflebeam, 2000, p. 311).

Using Stufflebeam’s (2001) approach, the evaluator engages a “representative stakeholder panel to help define the evaluation questions, shape evaluation plans, review draft reports and disseminate the findings” (p. 57). The stakeholder panel is the primary group with whom the evaluator interfaces regularly. The success of the evaluation can hinge upon these regular interactions because it is believed that without them, the evaluation approach will fail. The evaluator keeps the panel abreast of the formative information produced from the evaluation so that decisions about both the program and the evaluation can be made. Evaluations produce a comprehensive assessment of merit and program worth.

In sum, Stufflebeam’s (2001) evaluation approach engages stakeholders (usually in decision-making positions) in focusing the evaluation and in making sure the evaluation addresses their most important questions, providing timely, relevant information to assist decision making and producing an accountability record. Both formative and summative information become available to a panel of stakeholders that promotes use of the evaluation findings. By including multiple stakeholder perspectives, Stufflebeam increases the possibility that relevant value perspectives are represented, thus fostering a comprehensive evaluation of program value. With this as his evaluation focus, Stufflebeam is positioned as the first name on the use branch of the theory tree.

Malcolm Provus

Malcolm Provus was Director of Research and Evaluation in the Pittsburgh Public Schools. His views on evaluation were published in a 1971 book titled *Discrepancy Evaluation*. These views reflected a concern for providing evaluation information, particularly in a large district context with its broad organizational goals and complex political forces. Provus considers evaluation as a continuous process designed to assist program administrators. He specifies four required developmental stages and one optional stage. The four primary stages are *definition* (intended to specify the goals, processes, resources, etc.); *installation* (to identify discrepancies in the implementation of the program); *process* (to determine the extent of attainment of short-term outcomes or “enabling objectives”); and *product* (the extent of attainment of terminal or ultimate objectives). The fifth (optional) stage is *cost-benefit analysis*. One particular feature of *Discrepancy Evaluation* is its unique emphasis on program design (the definition stage). Also important to Provus is a strong interactive component involving program staff members and their reaching a working agreement (consensus) on the details of the design (Provus, 1971, p. 51).

Provus’s theoretical views (1971) strongly reflect the influence of Tyler. As with the objectives-oriented approaches, there is strong attention to the agreement on objectives, including the standard for judging those objectives. However, we have placed Provus on the use branch, paying particular heed to his emphasis on information to the management process. Provus notes that evaluation is the “handmaiden of administration in the management of program development through sound decision making” (p. 186).

Joseph Wholey

From Joseph Wholey’s academic training and exposure to large-scale program evaluations, one might anticipate his placement on the methods branch. Indeed, Wholey’s early writing depicts a theoretical view consistent with that of methods-oriented Suchman. However, Wholey’s long-standing participation in federal government programs made him sensitive to the needs of program managers. In many respects, Wholey’s focus on managers and policymakers (he is less concerned about stakeholders) is similar to Stufflebeam’s attention to decision makers. Much of Wholey’s (e.g., 1981, 1983) writings

have as a central focus the use of evaluation for the improvement of management. Wholey (1983) recognizes and attends in great detail to the complexity of managing an organization and views evaluation as an important process to “stimulate effective management.” Consequently, we have placed Wholey on the use branch.

In viewing an organization and its constraints, Wholey recognizes that obtaining evaluation information is costly. Thus, he proposes a four-stage procedure for the “sequential purchase of information.” The first of these, *evaluability assessment*, concerns itself with making an initial assessment (or preassessment) of the extent to which it is feasible to conduct an evaluation in terms of the organization and of the issues to be examined. Furthermore, the assessment considers the extent to which evaluation results are likely to be useful for program managers in effecting improved program performance.

The three other stages in the “sequential purchase of information” strategy are *rapid-feedback evaluation* (which focuses primarily on extant and easily collected information); *performance (or outcome) monitoring* (which measures program performance, usually in comparison to prior or expected performance); and *intensive evaluation* (which uses comparison or control groups to better estimate the effectiveness of program activities in causing observed results).

It is important to distinguish Wholey from others focused on large-scale programs whose writings also address the topic of evaluation use. In particular, it is important to compare Wholey to Weiss. Both have evaluation orientations that focus on large-scale (primarily federal government) programs, and each has an interest in and focuses on the policymaking context. Moreover, Weiss has played a very prominent role in defining the area of evaluation utilization and evoking a considerable amount of research in that area. Yet, for all her work on utilization, because of her emphasis on methods, Weiss is on the methods branch of this tree.⁶ Wholey is more firmly connected to the potential management benefits (particularly the short-term, instrumental, utilization) of his evaluation work.

Michael Patton

An extension of the use branch concerns itself with theorists who are not primarily focused on decision makers’ needs, but on emphasizing procedures

that would enhance the use of evaluation to a broader spectrum of identified stakeholders. Major research on evaluation utilization influenced this extension of the evaluation use branch.⁷ The view, based on this research, was that it is not sufficient to think of evaluation as necessarily related to decision makers, but it is necessary to think about the evaluator's obligation to try to help assure that utilization takes place. This point of view asserts that the evaluator should be proactive and not satisfied with an evaluation that might simply be put on the shelf. So, what should or could the evaluator do to make the evaluation more potentially utilizable?

The most prominent theoretical explication of the utilization (or use) extension was developed by Michael Patton (1978, 1986, 1997). At variance with earlier evaluation and decision-making theorists, Patton maintains that the evaluator must seek out individuals who are likely to be real users of the evaluation. Patton refers to them as "intended primary users." Indeed, Patton presents a flow chart of *utilization-focused evaluation* (UFE) in which the first step is the identification of intended users, including primary intended users. The other four major phases of UFE are users' commitment to the intended focus of the evaluation and to evaluation utilization; involvement in methods, design, and measurement; engagement, actively and directly interpreting findings and making judgments; and making decisions about further dissemination.

However, the essence of UFE derives from the major finding of research on evaluation utilization that a primary factor in obtaining use is the "personal factor" (Patton et al., 1977). The likelihood of an evaluation being utilized is greatly enhanced with the identification of people who have a stake in the evaluation and who personally care about the findings it generates. Thus, the strong focus on identifying primary intended users. Also flowing from the research finding on the personal factor is a concern for a commitment to utilization. This commitment should be on the part of an evaluator dedicated to and concerned about enhancing the possibility of utilization. Furthermore, the evaluator should be actively involved in developing intended users' commitment to potential utilization. The potential commitment is enhanced by engaging intended users actively and directly in all stages of the evaluation, fostering "buy-in."

The third major UFE concept emanating from the personal factor is the dictum that the evaluator be "active—reactive—adaptive." Evaluators need to be active in identifying intended users and focusing questions, reactive in continuing to learn about the evaluative situation, and adaptive "in altering

the evaluation questions and designs in light of their increased understanding of the situation and changing conditions” (Patton, 1997, p. 135). This utilization focus differs from the evaluation and decision-making thrust not only in its concern for identifying intended users but also in its willingness to be flexible in modifying prespecified evaluation questions and issues to better serve user needs.

In recent years, Patton has broadened the scope of UFE by incorporating other approaches. He has noted that “using evaluation to mobilize for social action, empower participants, and support social justice are options on the menu of evaluation process uses” (Patton, 1997, p. 103). Another addition is his introduction of the term *developmental evaluation*. In this activity, “the evaluator becomes part of the program design team or an organization’s management team, not apart from the team . . . but fully participating in decisions and facilitating discussion about how to evaluate whatever happens” (Patton, 1997, p. 106). Since in this role, the evaluator is, in essence, helping to develop the intervention, we believe that the evaluator has taken on a management-consulting role. By attaching these various new roles and modes to UFE, Patton perhaps clouds the clarity of the original utilization-focused evaluation approach.

Because of his work on evaluation utilization, Patton is placed on the use branch of the evaluation theory tree.

Marvin Alkin

In early years, Alkin was identified as focused on evaluation and decision-making concerns. Alkin’s early model (1972b) had many similarities to Stufflebeam’s CIPP model. The primary distinction between Alkin’s and Stufflebeam’s work is Alkin’s recognition that *process* and *product* have both summative and formative dimensions. Thus, one could look at process summatively (program documentation) or product formatively (meaning outcomes). The CSE model is used as the basis for the CSE evaluation kit, which has sold approximately 14,800 copies.

More recent writings (1991) place Alkin’s views as “user-oriented evaluation,” as he prefers to call it; to enhance the possibility of utilization, the focus is on the identified potential user(s). These views are more of a continuation of the original utilization-focused evaluation and less in keeping with

the new developmental evaluation direction that Patton's work has taken. In his theoretical works, Alkin strongly rejects the engagement of evaluators as valuing agents. Instead, he prefers to work with primary users at the outset of the evaluation process to establish value systems for judging potential outcome data. In interactive sessions, he presents a variety of simulated potential outcomes and seeks judgments (values) on the implications of each. Alkin acknowledges that there are conditions under which it is not tenable to engage intended primary users in this prejudgment process. Under such circumstances, he prefers to present evaluation data as factually as possible without imposing value judgment, unless there are extreme cases that demand that valuing take place.

We have placed Alkin on the use branch as an extension beyond the original Patton.

J. Bradley Cousins

The evaluation use limb continues to grow and has gained additional adherents who have made their own enhancements of these utilization ideas. Cousins and Earl's participatory evaluation (Cousins & Earl, 1992; Cousins & Whitmore, 1998) is a further extension of this branch. Cousins's presumption is that if we care about utilization, then the way to get utilization is to get buy-in. The way to get buy-in is to have program personnel participating in the evaluation.

Cousins's ideas flow from the concern for evaluation utilization expressed in the work of Patton. As such, Cousins continues to reflect the recognition of the importance of the personal factor in evaluation and the necessity for participation to heighten the possibility of utilization. However, he extends the notion of intended primary users to the necessity for organizing groups of intended users. Furthermore, his evaluations are designed for structured, continued, and active participation of these users, as opposed to Patton's user participation, which could take on a variety of different forms. Implicit within Cousins's approach is the understanding that utilization takes place within the context of an organization and is best accomplished as a part of organizational development. Cousins calls this "practical participatory evaluation" (Cousins & Earl, 1995).

Practical participatory evaluation is defined as "applied social research that involves trained evaluation personnel *and* practice-based decision

makers working in partnership” (Cousins & Earl, 1995, p. 8). Using this approach, primary users and evaluators are explicitly recognized as collaborators in the evaluation process. Practical participatory evaluation is grounded in an organizational learning framework that adopts strategies intended to enhance the learning capacity of organizations. Consequently, it is most suited for evaluation projects that “seek to understand programs with the expressed intention of informing and improving their implementation” (Cousins & Earl, 1995).

Cousins maintains that the desire for, and commitment to, program improvement attracts primary users to the evaluation process, that the partnership within the organization engages them, and that this in turn increases utilization. Responsibilities are shared equally between organization staff and the evaluator. The evaluator trains and supervises agency personnel in technical evaluation skills so that eventually they can coordinate new evaluation efforts. Thus, practitioners learn on the job. In time, the evaluator becomes a consultant responsible only for technical activities such as instrument design, data analysis, and technical reporting.

Thus, we have placed Cousins on the use branch of the tree. However, we have also noted from the research (Christie, 2003), that in pursuit of the goal of increasing utilization, Cousins has a preference for research methods; thus, the twig grows toward the methods branch.

Hallie Preskill

The work of Hallie Preskill continues on the use branch with a focus on organizational learning and development. She contends that substantial evaluation utilization can occur during the evaluation process and is a valuable tool for transformative learning. Influences on this work include Patton’s utilization-focused approach (particularly process use) and Cousins’s participatory evaluation model. As such, Preskill and her frequent coauthor Rosalie Torres are concerned with tailoring evaluations to fit the needs of primary users and getting the buy-in from program personnel participating in the evaluation.

Preskill’s theory is concerned with creating transformational learning within an organization through the evaluation process. *Transformational learning*, according to Preskill and Torres (2001), refers to a process where individuals, teams, and even organizations are identifying, examining, and understanding information needed to meet their goals. Transformational

learning is most likely to occur when “the purpose is to understand what others mean and to make ourselves understood” (p. 28). They maintain that to aid transformational learning, an evaluator should approach an organization (a) using a clinical approach, (b) spanning traditional boundaries between evaluator and program staff, and (c) diagnosing organizational capacity for learning.

These theorists advocate providing organizations with more than technical expertise (such as methods design and data analysis) to conduct an evaluation. Taking a clinical approach to evaluation allows for more focus on current needs, contextual settings, and historical surroundings of an organization. The approach “is inherently responsive to the needs of an organization and its members” (Preskill & Torres, 2001, p. 31) and allows the opportunity for reflection and for creating dialogue which in turn helps facilitate transformational learning.

To further improve transformational learning (and evaluation use), Preskill maintains that evaluators should span the traditional boundaries between evaluator and program staff. The evaluator should become a facilitator to help guide the learning process occurring during an evaluation. The facilitating process can be used to increase understanding of how the evaluation findings were arrived at, how they can be implemented, and what still needs to be done. The ultimate goal of this process is to create a seamless blend of program work, research, evaluation, and organizational development. Preskill views this approach as a way to help bridge the gap between what the program believes it’s doing and what it is actually doing.

The ability to diagnose an organization’s capacity for learning was stated as an important factor for learning and utilization (Preskill & Torres, 1998). Observing organizational culture, communication style, values, and leadership provides the evaluator with important information to assess the level of acceptance and use of evaluation results. The organizational diagnosis also offers evaluators a contextual setting to frame their findings and set realistic goals and expectations for evaluation use. Preskill contends that organizational learning can occur when results from the diagnosis are shared. This can provide an opportunity for reflection, transformational learning, and, ultimately, utilization.

Thus, Preskill has been placed on the utilization branch due to her focus on facilitating use during an evaluation and for her emphasis on the learning process that can occur when conducting an evaluation.

Jean King

Jean King extends the use branch with her focus on the applicability of participatory evaluation models. Influences on her work include Cousins and Earl's participatory approach and Patton's utilization-focused evaluation. King's work is distinguished from Patton's and Cousins's in two important ways. First, she is more overtly use orientated than Cousins. Second, she is more delineative than Patton in describing a procedure, *interactive evaluation practice* (IEP), for fostering participation and obtaining use.

King defines evaluation as "a process of systematic inquiry to provide sound information about the characteristics, activities, or outcomes of a program or policy for a valued purpose" (King & Stevahn, in press). This definition incorporates the importance of use by signifying that evaluations should be used for a "valued purpose." As such, King is concerned with tailoring evaluations in a collaborative manner with primary intended users to increase the chance of utilization.

King's theory is also concerned with creating a participatory environment throughout the evaluation process. King argues (1998) that for a participatory evaluation to succeed, there must be (a) an accepting power structure, (b) shared meaning of experiences amongst participants, (c) volunteers and leaders, (d) enough time, (e) enough resources, and (f) a great degree of interpersonal and organizational trust.

King advocates efforts of engagement in communication and discussion to create shared meaning amongst participants. She maintains that this process can lead to interest in formulating methods of inquiry, collecting, and interpreting data. King (1998) also urges evaluators to identify and foster leaders during the evaluation process. She contends that leaders are needed to "attract or recruit people to the evaluation process, who are eager to learn and facilitate the process . . . and who are willing to stay the course when things go wrong" (King, 1998, p. 64).

King focuses on trust building as one of the fundamental requirements for a successful participatory evaluation. She urges evaluators to pay close attention to the interpersonal dynamics occurring during evaluations (King & Stevahn, in press). She describes an evaluation without effective interpersonal interaction as "a machine without proper lubrication. At best, the machine will work. Gears may grind and teeth sheer off cogs. . . . At worst, the entire evaluation enterprise may grind to a screeching halt amidst a cloud of steam" (King & Stevahn, in press).

As a guiding principle, King provides the IEP framework. She defines interactive evaluation practice as “the intentional act of engaging people in making decisions, taking action, and reflecting while conducting an evaluation study” (King & Stevahn, in press). King maintains that for evaluators to practice IEP, they must focus on playing the role of decision maker, actor, and reflective practitioner.

King (King & Stevahn, in press) argues that as decision maker, the evaluator is compelled to make decisions “in order to move the evaluation process forward.” King describes the actor role of evaluators “as a performer in the theatrical event of the evaluation process.” In this role, the evaluator “plays a challenging role that is part leader, part manager, and part wise counselor.” As a reflective practitioner, the evaluator can help guide and assess evaluation progress. King contends that reflection “is especially important at key transition points in the study.”

The three roles suggested by King acknowledge the importance of the interpersonal factor when conducting evaluations. King’s suggested guidelines are aimed at increasing the chances of utilization and participatory evaluations, for, as King (1998) states, “You can lead people to evaluation, but you cannot make them participate” (p. 64). Thus, we have placed Jean King on the use branch.

John M. Owen

John Owen centers himself primarily in the domain of evaluation and organizational development and change. In this area, differences between organizational development consultants and evaluators are frequently debated. Evaluators and consultants are often distinguished from one another by the extent to which each participates in prescription rather than description. Traditionally, consultants provide prescriptions and recommendations, while evaluators provide descriptions and evaluative judgments. Owen maintains that the field of evaluation is changing, and, as such, the line between evaluating and consulting is blurring. He states,

“Increasingly, evaluators have adopted a more participatory framework for conducting evaluations that recognize the importance of representing stakeholder interest, and also seeking the investment of stakeholders in the construction of the evaluation process and product. Such collaborative arrangements clear the way for co-creating of prescriptions between evaluator and organization. (Owen & Lambert, 1998, p. 363)

This approach, Owen believes, supports a focus on evaluation utilization.

Most organizational development evaluators are concerned with issues related to use and stakeholder involvement. For Owen, evaluation utilization is a paramount consideration when planning, negotiating, designing, and conducting evaluations. Owen stresses the need for evaluators to be adaptable to stakeholder needs and concerns. It is this flexibility that distinguishes him from other organizational development evaluators, such as Preskill and Torres, who emphasize planned and purposeful evaluations. According to Owen, flexibility and a willingness to work collaboratively with stakeholders enhances the potential for an evaluation's use. He explains, "Evaluators who are prepared to adopt a more interactive and flexible stance can make a major contribution to the success of organizations in both private and public sectors" (Owen & Lambert, 1998, p. 364).

Owen supports stakeholder participation in multiple stages of the evaluation but believes that stakeholder involvement is most critical during the planning and negotiating stage. He states, "Evaluators must be prepared to acknowledge the interest framework of the clients, so that the knowledge they produce has salience to ensuing decisions about the evaluand under review" (Owen & Rogers, 1999, p. 65). Owen describes numerous techniques, guidelines, and frameworks for planning and negotiating flexible evaluations with stakeholders.

Owen has been placed on the use branch of the evaluation tree because of his emphasis on encouraging stakeholders to utilize evaluation findings. His specific location on the branch is as part of the cluster of organizational development evaluators, including Preskill and Torres, Cousins, and King.

David Fetterman

David Fetterman is the author of *Empowerment Evaluation* (1996), which he describes as a process that encourages self-determination among recipients of the program evaluation, often including "training, facilitation, advocacy, illumination and liberation." Pursuing the empowerment evaluation goal, which is to foster self-determination rather than dependency, program participants—including clients—essentially conduct their own evaluations. The outside evaluator often serves as a coach or additional facilitator, providing clients with the knowledge and tools for continuous self-assessment and accountability. Fetterman feels that training participants to evaluate their own

programs and coaching them through the design of an evaluation is an effective form of empowerment.

Fetterman (1994) describes two general forms of empowerment evaluation that are only subtly different. The main distinction between the two is the extent to which the evaluator participates in the evaluation process. In the first case, evaluators teach program participants to conduct their own program evaluations, making them more self-sufficient. In the second case, the evaluator serves as a coach to help others conduct their own evaluations. Using this approach, evaluators allow participants to shape the direction of the evaluation, suggest ideal solutions to their problems, and then take an active role in making social change. Fetterman sees all empowerment evaluators as having the potential to serve as “illuminating and liberating facilitators, assisting program participants, freeing themselves from traditional roles and expectations” (p. 306).

For Fetterman (1998), the end point of the evaluation is not the assessment of the program’s worth. In his view, value and worth are not static. He sees evaluation as an ongoing process: “Through the internalization and institutionalization of self-evaluation processes and practices, a dynamic and responsive approach to evaluation can be developed to accommodate shifts in populations, goals, value assessments and external forces” (p. 382).

It is important to distinguish between participatory and empowerment evaluation. Although these approaches employ similar practices, to consider them synonymous would be a fundamental misunderstanding of the theories. It is primarily the goal rather than the practice of the evaluation that differentiates the two. In practice, the conception of stakeholder involvement that underlies both participatory and empowerment evaluation looks very similar; that is, stakeholders decide upon and assist in conducting all aspects of the evaluation, including design, implementation, analysis, and interpretation. But since participatory evaluation emerges from a utilization framework, the goal of the participatory process is increased utilization through these activities, as opposed to empowering those that have been oppressed, which is political or emancipatory in nature.

A FINAL NOTE

There are two main challenges we faced in writing this chapter. On the one hand, in sorting out the views and positions of theorists, we needed to make a specific placement on a particular branch of the tree. The second challenge was determining which theorists to include on the tree.

Let us consider the first of these issues. The theory tree is posited on the view that ultimately, one or another of the three dimensions, depicted as branches, was of the highest priority for each theorist. We did not have the luxury of allocating percentages to each branch, as was suggested to us by one theorist at a recent meeting of the American Evaluation Association. To do so would create a state of confusion and fail to show the relationship between theorists in a clean-cut manner.

An example will demonstrate the way that decisions were made and the difficulty of making these judgments.

Over a year ago, one of our students, as part of a class assignment, contacted David Fetterman to inquire about the influences on his evaluation theoretic positions. He cited three: (1) anthropological/ethnographic (Spindler and Pelto being the two most prominent), (2) social justice (House being the most prominent), and (3) evaluation utilization (Patton and Alkin being the two most prominent). Now, where to place him?

An emphasis on educational anthropology/ethnography centers on methods. Such a focal point would mean that the most important element of evaluation would be the use of ethnographic methods. And although Fetterman has a strong preference for using qualitative methods in evaluation, he certainly does not limit his work to the use of such methods. An emphasis on social justice would imply that the purpose of the evaluation is to place a value on the outcome of the evaluation that in some way supports social justice ideals. In the case of Fetterman, social justice is served through the elimination of the reliance on others to engage in activities that lead to program improvement; it does not mean judging a program based on whether or not it meets the criteria derived from a specific set of socially just ideals. An emphasis on utilization implies that the main concern is that utilization of evaluation findings occurs. In the case of Fetterman, the utilization is process use, namely, that the process of conducting the evaluation leads to empowerment.

It may be that Fetterman was led to evaluation through his work in educational anthropology and ethnography (see Fetterman, 1984, 1988; Fetterman & Pitman, 1986). In his earliest writings, Fetterman identified participant observation as a key element of ethnographic educational evaluation and stressed the importance of using a cultural perspective to interpret and analyze data. He advocated strongly for data interpretation through the eyes of those involved in the evaluation rather than through the eyes of the researcher. Yet despite these anthropological roots, he began to think of the principal

focus of an evaluation in another way; that is, ethnography became the primary methodology for the conduct of evaluations that had other primary motivations.

Another question arises in analyzing Fetterman's work: Is the primary focus valuing outcomes in a way that achieve social justice, which would place him on the valuing branch? Or, alternatively, is the primary focus in Fetterman related to the use, specifically the process use, of evaluation? We believe that, as Fetterman describes it, the act of empowering focuses on the process of engaging in evaluation; that is, by training individuals to engage in evaluation activities, they become empowered. In the language of evaluation utilization, empowerment evaluation involves instrumental process use. Thus, while noting a deep concern for social justice and a strong preference for (and early evaluation roots in) anthropological/ethnographic methods, we were led to place Fetterman on the utilization branch.

The determination of a theorist's placement on a branch of the evaluation theory tree was not always this difficult, but it always required careful consideration and the analysis of trade-offs.

The second issue, which evaluation theories to include on the tree, also offered a particular challenge. This has been addressed to a certain extent earlier in this chapter by designating distinctions between evaluation methodologists, evaluation analysts, evaluation writers, and evaluation theorists. Furthermore, this book represents the theoretical perspectives of the English-speaking world. Clearly, there may be broad theoretical views written in other languages and not translated into English. However, it is important to note that many of the theorists in this book have had their works translated into other languages and have had substantial worldwide influence.

A further limitation on theory selection is that we have focused on general evaluation theories, as opposed to works that specifically focus on evaluation in a particular field (e.g., education, social welfare, and public health). Thus, evaluation works such as Steckler and Linnan (2002) and Astin (1991) have not been included.

Finally, theories included were able to be classified onto a single branch of the tree. A number of the theorists offered difficulty, but we were ultimately able to make a determination of an appropriate branch placement (e.g., Fetterman, discussed above). Whether some theories were not included due to their comprehensiveness or our conceptual inability is unclear. A particular example comes to mind: the work of Mark et al. (2000). These authors view social

betterment as the ultimate objective of evaluation and present a point of view grounded in what they refer to as a “common sense realist philosophy.” Furthermore, Mark et al. maintain that the evaluation to be conducted is dependent upon the particular purpose to be served. They provide four purposes: assessing merit and worth, program and organizational improvement, knowledge development, and oversight and compliance. The first three of these correspond reasonably to the three branches of the tree, and the fourth represents the accountability root, which we maintain pervades the various theories. As Jennifer Greene notes in commenting on their approach, “Evaluators of many theoretical and practical persuasions can find a comfortable place for themselves within the big tent that is evaluation as assisted sense making. And, they need not worry that the tent will collapse” (Greene, book jacket quotation on Mark et al., 2000). The very diversified nature of this perspective, while a great strength in presenting an understanding of evaluation, precludes its inclusion on the tree.

NOTES

1. This additional dimension of his work is reflected, in part, in his work with Cook (Cook & Campbell, 1976, 1979, 1986).

2. In actuality, it is hard to designate beginning points in the theory of evaluation. An examination of Suchman’s work shows extensive reliance on a number of evaluation writers, particularly the “Memorandum on Program Evaluation” by Riecken (1953/1972).

3. In the Berk et al. (1985) manuscript, randomized field experiment approach is referred to as “social policy experimentation.”

4. Cronbach is discussed later in this chapter.

5. Scriven is discussed later in this chapter.

6. Weiss’s work is more fully discussed in an earlier section of this chapter.

7. Weiss did some of the early writing on evaluation utilization (e.g. Weiss, 1972b). Subsequent research was conducted by Alkin, Kosecoff, Fitz-Gibbon and Seligman (1974); Patton (1978); Alkin, Daillak, and White (1979); King and Pechman (1982); Braskamp, Brown, and Newman (1982); and others.

REFERENCES

- Alkin, M. (1972a). Accountability defined. *Evaluation Comment: The Journal of Educational Evaluation*, 3, 1-5.
- Alkin, M. (1972b). Evaluation theory development. In C. Weiss (Ed.), *Evaluation action programs* (pp. 105-117). Boston: Allyn & Bacon.

- Alkin, M. (1990). *Debates on evaluation*. Newbury Park, CA: Sage.
- Alkin, M. (1991). Evaluation theory development. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter-century (90th yearbook of the National Society for the Study of Education, Part II)* (pp. 91-112). Chicago: University of Chicago Press.
- Alkin, M., Daillak, R., & White, P. (1979). *Using evaluations: Does evaluation make a difference?* (Sage Library of Social Research, Vol. 76). Beverly Hills, CA: Sage.
- Alkin, M., Hofstetter, C., & Ai, X. (1998). Stakeholder involvement in evaluation. In A. Reynolds & H. Walberg (Eds.), *Evaluation for educational productivity*. Greenwich, CT: JAI.
- Alkin, M., Kosecoff, J., Fitz-Gibbon, C., & Seligman, R. (1974). Evaluation and decision making: The title VII experience. *CSE Monograph, 4*.
- Astin, A. (1991). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. New York: Macmillan.
- Berk, R., Boruch, R., Chambers, O., Rossi, P., & Witte, O. (1985). Social policy experimentation: A position paper. *Evaluation Review, 9*(4), 387-429.
- Bloom, B., Englehart, D., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: Handbook I. The cognitive domain*. New York: David McKay.
- Boruch, R. (1997). *Randomized experiments for planning and evaluation*. Thousand Oaks, CA: Sage.
- Boruch, R., McSweeney, A., & Soderstrom, E. (1978). Randomized field experiments for program planning, development and evaluation. *Evaluation Quarterly, 2*(4), 655-695.
- Boruch, R., Synder, B., & DeMoya, D. (2000). The importance of randomized field trials. *Crime & Delinquency, 46*(2), 156-180.
- Braskamp, L., Brown, R., & Newman, D. (1982). Studying evaluation utilization through simulations. *Evaluation Review, 6*(1), 114-126.
- Campbell, D. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54*, 297-312.
- Campbell, D. (1975a). Assessing the impact of planned social change. In G. M. Lyons (Ed.), *Social research and public policies* (pp. 3-45). Hanover, NH: Dartmouth College, Public Affairs Center.
- Campbell, D. (1975b). "Degrees of freedom" and the case study. *Comparative Political Studies, 8*, 1178-1193.
- Campbell, D., & Stanley J. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Chen, H., & Rossi, P. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*, 283-302.
- Chen, H., & Rossi, P. (1987). The theory-driven approach to validity. *Evaluation and Program Planning, 10*, 95-103.
- Christie, C. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation, 97*, 7-35.

- Cook, T., & Campbell, D. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 223-326). Chicago: Rand McNally.
- Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Cook, T., & Campbell, D. (1986). The causal assumptions of quasi-experimental practice. *Synthese*, 68, 141-180.
- Cook, T., & Gruder, C. (1978). Metaevaluation research. *Evaluation Quarterly*, 2, 5-51.
- Cousins, J., & Earl, L. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14(4), 397-418.
- Cousins, J., & Earl, L. (Eds.). (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London: Falmer.
- Cousins, J., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, 80, 5-23.
- Cronbach, L. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Cronbach, L., & Associates (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco: Jossey-Bass.
- Cronbach, L., & Snow, R. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.
- Durkheim, E. (1966). *The rules of sociological method* (8th ed.). George C. Catlin (Ed.). (Solovay & John H. Mueller, Trans.). New York: Free Press.
- Eisner, E. (1985). *The art of educational evaluation: A personal view*. Philadelphia, PA: The Falmer Press.
- Eisner, E. (1976). Educational connoisseurship and criticism: Their form and function in educational evaluation. *Journal of Aesthetic Evaluation or Education*, 10, 135-150.
- Eisner, E. (1994). *The educational imagination: On the design and evaluation of educational programs* (3rd ed.). New York: Macmillan.
- Eisner, E. (1998). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. Upper Saddle River, NJ: Merrill.
- Eisner, E. (1991a). *The enlightened eye*. New York: Macmillan.
- Eisner, E. (1991b). Taking a second look: Educational connoisseurship revisited. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter-century* (90th yearbook of the National Society for the Study of Education, Part II) (pp. 169-187). Chicago: University of Chicago Press.
- Fetterman, D. (1984). Ethnography in educational research: The dynamics of diffusion. In D. M. Fetterman (Ed.), *Ethnography in educational evaluation* (pp. 21-35). Beverly Hills, CA: Sage.
- Fetterman, D. (1988). A national ethnographic evaluation: An executive summary of the ethnographic component of the Career Intern Program Study. In D. M. Fetterman (Ed.), *Qualitative approaches to evaluation in education: The silent scientific revolution* (pp. 262-273). New York: Praeger.

- Fetterman, D. (1994). Steps of empowerment education: From California to Cape Town. *Evaluation and Program Planning*, 17(3), 305-313.
- Fetterman, D. (1996). Empowerment evaluation: An introduction to theory and practice. In D. M. Fetterman, S. J. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Thousand Oaks, CA: Sage.
- Fetterman, D. (1998). Empowerment evaluation and accreditation in higher education. In E. Chelmsky & W. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 381-395). Thousand Oaks, CA: Sage.
- Fetterman, D., & Pitman, M. (Eds.). (1986). *Educational evaluation: Ethnography in theory, practice, and politics*. Beverly Hills, CA: Sage.
- Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.
- Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Hamilton et al. (Eds.). (1977). *Beyond the numbers game: A reader in educational evaluation*. D. Hamilton, D. Jenkins, C. King, & M. Parlett (Eds.). New York: Macmillan Education. (Also 1978, San Francisco: McCutcheon)
- Hammond, R. (1973). Evaluation at the local level. In B. R. Worthen & J. R. Sanders (Eds.), *Educational evaluation: Theory and practice*. Belmont, CA: Wadsworth.
- House, E. (1991). Evaluation and social justice: Where are we? In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (90th yearbook of the National Society for the Study of Education, Part II) (pp. 233-247). Chicago: University of Chicago Press.
- House, E. (1993). *Professional evaluation: Social impact and political consequences*. Newbury Park, CA: Sage.
- House, E. (2001). Unfinished business: Causes and values. *The American Journal of Evaluation*, 22, 309-315.
- Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage.
- King, J. A. (1998, Winter). Making sense of participatory evaluation on practice. *New Directions in Evaluation*, 80, 57-67.
- King, J. A., & Pechman, E. (1982). *Improving evaluation use in local schools*. Washington, DC: National Institute of Education.
- King, J. A., & Stevahn, L. (in press). *Interactive evaluation in practice: Managing the interpersonal dynamics of program evaluation*. Thousand Oaks, CA: Sage.
- Krathwohl, D., Bloom, B., & Masia, B. (1964). *Taxonomy of educational objectives. Handbook II: The affective domain*. New York: David McKay.
- Kuhn, T. (1962). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.
- MacDonald, B. (with Jenkins, F., Kemmis, S., & Verma, G.). (1979). In G. K. Verma & C. Bugley (Eds.), *Racism and educational evaluation education and identity*. London: Macmillan.
- Madaus, G., & Stufflebeam, D. (1989). *Educational evaluation: Classic works of Ralph W. Tyler*. Boston: Kluwer.

- Mark, M., Henry, G., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving public and nonprofit policies and programs*. San Francisco: Jossey-Bass.
- Marx, K. (1932). *Capital: A critique of political economy. Vol. 1*. Frederick Engels (Ed.). (Samuel Moore & Edward Aveling, Trans.). Revised and amplified according to the fourth German Edition by Ernest Untermann. Chicago: Charles H. Kerr.
- Metfessel, N., & Michael, W. (1967). A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. *Educational and Psychological Measurement*, 27, 931-943.
- Northwest Regional Educational Laboratory. (1977, January). *3-on-2 evaluation report, 1976-1977* (Vols. I, II, & III). Portland, OR: Author.
- Owen, J., & Lambert, F. (1998). Evaluation and the information needs of organizational leaders. *American Journal of Evaluation*, 19, 355-365.
- Owen, J., & Rogers, P. (1999). *Program evaluation: Forms and approaches*. London: Sage.
- Owens, T. (1973). Education evaluation by adversary proceeding. In E. R. House (Ed.), *School evaluation: The politics and process*. Berkeley, CA: McCutcheon.
- Patton, M. Q. (1978). *Utilization-focused evaluation*. Beverly Hills, CA: Sage.
- Patton, M. Q. (1986). *Utilization-focused evaluation* (2nd ed.). Beverly Hills, CA: Sage.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q., Grimes, P., Guthrie, K., Brennan, N., French, B., & Blyth, D. (1977). In search of impact. In C. H. Weiss (Ed.), *Using social research in public policy making* (pp. 141-164). Lexington, MA: Lexington Books.
- Popham, W. (1973). *Evaluating instruction*. Englewood Cliffs, NJ: Prentice Hall.
- Popham, W. (1975). *Educational evaluation*. Englewood Cliffs, NJ: Prentice Hall.
- Popham, W. (1988). *Educational evaluation* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Popham, W., & Carlson, D. (1977). Deep dark deficits of the adversary evaluation model. *Educational Researcher*, 6(6), 3-6.
- Popper, Karl R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. London: Routledge.
- Preskill, H., & Torres, R. T. (1998, November 4-7). *Evaluative inquiry as transformative learning*. Panel/paper presented at the American Evaluation Association Annual Conference, Chicago.
- Preskill, H., & Torres, R. T. (Eds.). (2001). The learning dimension of evaluation use. *New Directions for Evaluation* (Vol. 88). San Francisco: Jossey-Bass.
- Provus, M. (1971). *Discrepancy evaluation*. Berkeley, CA: McCutcheon.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
- Riecken, H. (1972). Memorandum on program evaluation. In Carol Weiss (Ed.), *Evaluating action programs: Readings in social action and education*. Boston: Allyn & Bacon. (Original work published 1953)

- Rossi, P., & Freeman, H. (1985). *Evaluation: A systematic approach* (3rd ed.). Beverly Hills, CA: Sage.
- Rossi, P., Freeman, H., & Lipsey, M. (1999). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.
- Rossi, P., Freeman, H., & Wright, S. (1979). *Evaluation: A systematic approach*. Beverly Hills, CA: Sage.
- Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum evaluation. American Educational Research Association Monograph Series on Evaluation No. 1*. Chicago: Rand McNally.
- Scriven, M. (1972a). The methodology of evaluation. *Perspectives of curriculum evaluation*. Chicago: Rand McNally.
- Scriven, M. (1972b). Prose and cons about goal-free evaluation. *Evaluation Comment: The Journal of Educational Evaluation*, 3(4), 1-7. UCLA, Center for the Study of Evaluation.
- Scriven, M. (1974). Evaluation perspectives and procedures. In J. W. Popham (Ed.), *Evaluation in education: Current application* (pp. 3-93). Berkeley, CA: McCutcheon.
- Scriven, M. (1976). Evaluation bias and its control. In G. V. Glass (Ed.), *Evaluation studies review annual* (Vol. 1). Beverly Hills, CA: Sage.
- Scriven, M. (1983). Evaluation ideologies. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 229-260). Boston: Kluwer-Nijhoff.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.
- Scriven, M. S. (1986). New frontiers of evaluation. *Evaluation Practice*, 7, 7-44.
- Shadish, W., Cook, T., & Leviton, L. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
- Stake, R. (1975). *Program evaluation, particularly responsive evaluation* (Occasional Paper No. 5). Kalamazoo: Western Michigan University Evaluation Center.
- Stake, R. (2000). A modest commitment to the promotion of democracy. *New Directions for Evaluation*, 85, 97-107.
- Stake, R. (2001). A problematic heading. *The American Journal of Evaluation*, 22, 349-354.
- Steckler, A., & Linnan, L. (2002). *Process evaluation for public health interventions and research*. San Francisco, CA: Jossey-Bass.
- Stufflebeam, D. (1983). The CIPP model for program evaluation. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 117-141). Boston: Kluwer-Nijhoff.
- Stufflebeam, D. (2000). Lessons in contracting for evaluations. *The American Journal of Evaluation*, 21, 293-314.
- Stufflebeam, D. (2001). Interdisciplinary Ph.D. programming in evaluation. *The American Journal of Evaluation*, 22, 445-455.

- Suchman, E. (1967). *Evaluative research: Principles and practice in public service and social action programs*. New York: Russell Sage.
- Tineke, A., & Stake, R. (2001). Stakes responsive evaluation: Core ideas and evolution. *New Directions for Evaluation*, 92, 7-20.
- Tyler, R. W. (1942). General statement on evaluation. *Journal of Educational Research*, 35, 492-501.
- Wagner, R. B. (1989). *Accountability in education: A philosophical inquiry*. New York: Routledge.
- Weiss, C. (1972a). *Evaluation research: Methods of assessing program effectiveness*. Englewood Cliffs, NJ: Prentice Hall.
- Weiss, C. (Ed.). (1972b). *Evaluating action programs: Readings in social action and education*. Boston: Allyn & Bacon.
- Weiss, C. (1973). Where politics and evaluation research meet. *Evaluation*, 1, 37-45.
- Weiss, C. (1976). Using research in the policy process: Potential and constraints. *Policy Studies Journal*, 4, 224-228.
- Weiss, C. (1981). Doing research or doing policy [Review of *Toward reform of program evaluation*]. *Evaluation and Program Planning*, 4, 397-402.
- Weiss, C. (1991). Evaluation research in the political context: Sixteen years and four administrations later. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (90th yearbook of the National Society for the Study of Education, Part II) (pp. 211-231). Chicago: University of Chicago Press.
- Weiss, C., & Bucuvalas, M. (1980). *Social science research and decision-making*. New York: Russell Sage.
- Wholey, J. (1981). Using evaluation to improve program performance. In R. A. Levine, M. A. Solomon, G. M. Hellstern, & H. Wollman (Eds.), *Evaluation research and practice: Comparative and international perspectives* (pp. 92-106). Beverly Hills, CA: Sage.
- Wholey, J. (1983). *Evaluation and effective public management*. Boston: Little, Brown.
- Wolf, R. L. (1975). Trial by jury: A new evaluative method. *Phi Delta Kappan* 57(3), 185-187.
- Wolf, R. L. (1979). The use of judicial evaluation methods in the formulation of educational policy. *Educational Evaluation and Policy Analysis*, 1(93), 19-28.
- Worthen, B., & Sanders, J. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. White Plains, NY: Longman.