

Qualitative Program Evaluation

Practice and Promise

JENNIFER C. GREENE

Sylvia Winslow, chair of a state legislative committee on health, is concerned about the effectiveness of the state's prenatal care services. Recent reports from varied sources—most dramatically in the media—have underscored the difficulties that rural women especially have in finding prenatal care and the tragedies that result from inadequate care. Sylvia asks the staff of her health committee for a comprehensive update on the state's prenatal care program, with special focus on issues of access and quality.

Albert Peters enrolls his son in a new program at the local high school that claims to be offering a different vision of secondary education. Critical elements of this vision for Albert and his son include heterogeneous grouping, a curriculum infused with multicultural ideas and values, and a strong mentoring system in which faculty guide each youth's program and progress. Albert genuinely believes that a program, such as this one might be, in which all kids are respected and assumed to have unique and valuable gifts is the only way his son will make it through high school. It is October already, and Albert wants some information on the nature and quality of this educational experience for his son. Albert calls the high school principal to request such information.

Chapter 1, the U.S. federal government's long-term compensatory education program, is up for reauthorization in two years. In anticipation of the probable debate, the chair of the Senate Education and Labor Committee asks the U.S. General Accounting Office to provide a summary assessment of Chapter 1 success over the past five years.

Evan Gonzalez, the human resource director at IVM, a major electronics corporation, is alarmed at the increasing number of employees utilizing the company's counseling and psychological services. Evan's concern is both for the employees: What is going on in the company or in specific communities to invoke such a demand for mental health services?—and for the company, as these services are extremely costly. Evan raises his questions at the next management meeting.

Sam Brown has been advocating for the homeless in his county for five years now, and he is frustrated. He needs some kind of catalyst to mobilize his constituents and to energize concerned citizens toward concrete action by the county board. Sam could have current information on the number of homeless in the county, on their personal histories, and on the social costs of homelessness. Well, that might just do the trick.

AUTHOR'S NOTE: My sincere thanks to the reviewers and the editors for their constructive contributions to this chapter.

Contexts of Program Evaluation

The above are examples of some of the many scenarios of social program evaluation. They vary in the nature of the social issues involved, in the perspectives taken on the issues, in the geographic scope of services to be reviewed, in the kinds of information sought, and in the stated purposes for which the information will be used. Underlying these differences, however, are some fundamental commonalities that demarcate evaluation contexts, and that thereby distinguish program evaluation as a unique form of social inquiry.

Perhaps most distinctive about program evaluation is its political inherency (Patton, 1987), the "recognition that politics and science are both integral aspects of evaluation" (Cronbach & Associates, 1980, p. 35), and that "the evaluator has political influence even when he does not aspire to it" (Cronbach & Associates, 1980, p. 3). Evaluations are conducted on social programs—most important, on social programs in the public domain.¹ Social programs are manifest responses to priority individual and community needs and are themselves "the creatures of political decisions. They [are] proposed, defined, debated, enacted, and funded through political processes, and in implementation they remain subject to [political] pressures—both supportive and hostile" (Weiss, 1987, p. 47). So program evaluation is integrally intertwined with political decision making about societal priorities, resource allocation, and power. "By its very nature [evaluation] makes implicit political statements about such issues as the problematic nature of some programs and the unchallengeability of others" (Weiss, 1987, p. 48).

Moreover, the work of social program evaluators is framed by the concerns and interests of selected members of the setting being evaluated. Evaluation questions about the significance of program goals or about the quality and effectiveness of program strategies reflect not inquirer autonomy or theoretical predictions, but rather a politicized process of priority setting. In all evaluation contexts there are multiple, often competing, potential audiences—groups and individuals who have vested interests in the program being evaluated, called *stakeholders* in evaluation jargon. These range from policy makers and funders like Sylvia Winslow and the U.S. Congress to program administrators and staff like Evan Gonzalez of IVM, to intended beneficiaries like Albert Peters and homeless advocate Sam [redacted], and to the citizenry at large. And so, unlike other social scientists, who assume an audience of peers/scholars, evaluators must negotiate questions will be addressed and whose interests will be served by their work.

Evaluation results then enter the political arena of program and policy decision making not

as decontextualized, abstract, or theoretical knowledge claims, but rather as practical knowledge claims, as empirically justified value judgments about the merit or worth of the program evaluated. Evaluators describe and infer about practical matters, about the significance of concrete program experiences for various stakeholders. But evaluators do more than describe and infer. At root, evaluation is about valuing (Scriven, 1967) and judging (Stake, 1967). Hence evaluators also infuse directly into the political strands of social policy making the standards or criteria used for rendering judgments. Like the selection of evaluation questions and audiences, determining the standards against which a program will be judged is a contested task. Increasingly, particularly in qualitative evaluations, these valuing standards are identified and offered pluralistically, as multiple sets. Program effectiveness, for example, has many hues, depending on one's vantage point in both space and time. Administrators might well understand effectiveness as efficiency, beneficiaries as significant relief from life's daily struggles, and funders as the long-term realization of tax dollars saved.

Evaluation Methodologies

Yet neither these diverse criteria for program effectiveness nor different stakeholders' widely divergent evaluation questions can be equally well addressed by the same evaluation methodology. In this respect, it is the fundamental political nature of program evaluation contexts, intertwined with the predispositions and beliefs of the evaluator, that shape the contours of evaluation methodologies and guide the selection of a specific evaluation approach for a given context. Different evaluation methodologies are expressly oriented around the information needs of different audiences—from the macro program- and cost-effectiveness questions of policy makers to the micro questions of meaning for individual participants. These varied audience orientations further represent, explicitly or implicitly, the promotion of different values and political stances. Evaluation methodologies hence constitute coordinated frameworks of philosophical assumptions (about the world, human nature, knowledge, ethics), integrated with ideological views about the role and purpose of social inquiry in social policy and program decision making, with accompanying value stances regarding the desired ends of programs and of inquiry, and finally—last as well as least—with complementary methods preferences. Again, it is because evaluation is politically contextualized that constitutive differences in evaluation

TABLE 33.1 Major Approaches to Program Evaluation

<i>Philosophical Framework</i>	<i>Ideological Framework/ Key Values Promoted</i>	<i>Key Audiences</i>	<i>Preferred Methods</i>	<i>Typical Evaluation Questions</i>
Postpositivism	Systems theory/ efficiency, accountability, theoretical causal knowledge	High-level policy and decision makers	Quantitative: experiments and quasi-experiments, systems analysis, causal modeling, cost-benefit analysis	Are desired outcomes attained and attributable to the program? Is this program the most efficient alternative?
Pragmatism	Management/ practicality, quality control, utility	Mid-level program managers, administrators, and other decision makers	Eclectic, mixed: structured and unstructured surveys, questionnaires, interviews, observations	Which parts of the program work well and which need improvement? How effective is the program with respect to the organization's goals? With respect to beneficiaries' needs?
Interpretivism	Pluralism/ understanding, diversity, solidarity	Program directors, staff, and beneficiaries	Qualitative: case studies, interviews, observations, document review	How is the program experienced by various stakeholders?
Critical, normative science	Emancipation/ empowerment, social change	Program beneficiaries, their communities, other "powerless" groups	Participatory: stakeholder participation in varied structured and unstructured, quantitative and qualitative designs and methods; historical analysis, social criticism	In what ways are the premises, goals, or activities of the program serving to maintain power and resources inequities in the society?

methodologies extend well beyond alternative methods and also beyond alternative philosophies of science (see Guba, 1990) to incorporate alternative ideologies (Scriven, 1983) and alternative philosophies of ethics, democracy, and justice (House, 1980).

Table 33.1 offers a descriptive categorization of four major genres of evaluation methodologies. The first, which represents the historically dominant tradition in program evaluation, is oriented around the macro policy issues of program effectiveness and cost efficiency. Sylvia Winslow's questions about the quality of her state's prenatal health care services and the U.S. Senate's questions about Chapter 1 success illustrate the broad program effectiveness issues this evaluation genre is designed to address. In this genre primary emphasis is placed on program effectiveness as outcomes and, concomitantly, on the social value of accountability. Early postpositivist evaluation was typified by large-scale studies of Great Society programs, such as the Head Start evaluation (Cicirelli

& Associates, 1969) and, later, the New Negative Income Tax Experiment (Rossi & 1978). These studies demonstrated well the use of experimental logic to meet the demands of evaluation settings. Even so, with "old castles unthroned, but not abolished" (Cook, 1983) and new efforts to reclaim the primacy of science and hence of scientists, in social program evaluation as best represented by theory-driven evaluation (Chen & Rossi, 1983)—postpositivist evaluation retains a strong position amidst theoretical evaluation practitioners and a still-dominant position among evaluation audiences. *How can this be so? How can it be so if it doesn't have a control group?* is a familiar lament from the field.

The second genre of evaluation methodologies arose largely in response to the demands of mental science to provide timely information for program decision making. The questions of these methodologies are those of decision making and hence to

primary emphasis on producing useful information, their practical and pragmatic value base, and their eclectic methodological stance. Evaluators in this genre pragmatically select their methods to match the practical problem at hand, rather than as dictated by some abstract set of philosophical tenets (Howe, 1988; Patton, 1988). For example, in order to decide what action if any to take in the face of IVM employees' increasing use of psychological services, Evan Gonzalez and other IVM managers are likely to need a variety of qualitative, quantitative, and perhaps even historical information related to the experiences and contexts of the workplace. Decision- and utilization-oriented evaluators focus on providing support for efficient and effective program management. As an unwavering champion of practical, utilization-focused evaluation, Michael Patton's (1990) approach to qualitative evaluation clearly falls within this genre.

Yet it is in the third cluster that more traditionally qualitative approaches to evaluation have found their home. Part of the interpretive turn in social science, these approaches share a common grounding in a basically interpretive philosophy of science (Smith, 1989; Soltis, 1990), a value orientation that characteristically promotes pluralism in evaluation contexts, and a case study methodological orientation with an accompanying reliance on qualitative methods. Part of the responsive tradition in program evaluation, these approaches seek to enhance contextualized program understanding for stakeholders closest to the program (like Albert Peters and his son), and thereby promote values of pluralism as well as forge direct channels to program improvement. Robert Stake's (1975) and Egon Guba and Yvonna Lincoln's (1981) responsive approaches to evaluation, and, though not as good a fit, Elliot Eisner's (1976) connoisseurship evaluation are major exemplars of qualitative evaluation approaches in this genre.

Finally, the fourth cluster represents the more recent *normative turn* in social science. The feminist, neo-Marxist, critical, and other theorists in this genre promote "openly ideological" forms of inquiry that seek to illuminate the historical, structural, and value bases of social phenomena and, doing so, to catalyze political and social change and greater justice, equity, and democracy. A qualitative evaluation stance well matches home-advocate Sam Brown's need for information that would help catalyze action toward greater justice for homeless people. Although many proponents of a normative approach to evaluation are rhetorically strong and short on guidelines for the practice of normative evaluation, some work on the practice of normative evaluation is being done (Greene, 1991; McTaggart, 1990; Sirotnik & Oakes, 1990). For the present discussion, the democratic evaluation approach advanced by British and Australian practitioners (Guba & Lincoln, 1976) is an important exemplar. As

well, Guba and Lincoln's (1989) more recent development of fourth-generation evaluation bears examination as it promotes an activist ideology while maintaining a grounding in an essentially interpretivist philosophy.

Explicit recognition of the ideological contours of program evaluation did not always exist. Rather, both inside and outside the field, a methods orientation has predominated. One continuing legacy of this orientation is the naming of different evaluation approaches by their primary methods, no more common than in the label of "qualitative evaluation" for any approach that utilizes primarily qualitative methods. Yet, as just argued, what importantly distinguishes one evaluation methodology from another is not methods, but rather whose questions are addressed and which values are promoted. Among the major extant evaluation approaches that rely on qualitative methods, there is only some consensus on these political and value dimensions. Patton, Stake, MacDonald, and Guba and Lincoln represent distinct, even competing, positions on these dimensions. Qualitative research traditions show similar variability (Atkinson, Delamont, & Hammersley, 1988; Jacob, 1987).

In my discussion in this chapter I will emphasize distinctions among significant evaluation approaches that incorporate qualitative methods. And that's progress! Not long ago, interpretivist philosophies and qualitative methods were just gaining a foothold in the evaluation community, amidst considerable, often acrimonious, debate. In the next section, I offer a brief historical perspective on the ascendance and acceptance of qualitative methods in program evaluation methodologies. This is followed by an elaboration of the philosophical bases of qualitative evaluations, and then an examination of qualitative evaluations in practice. I conclude the chapter by noting the continuing challenges for program evaluators and providing a summary assessment regarding the future promise of qualitative evaluations. I will argue that because of their inherent paradigmatic relativity, many qualitative evaluation approaches can effectively respond to or be shaped to fit diverse and emergent inquiry forms and functions. Yet, absent a coherent value orientation or vision (such as the social activism that frames fourth-generation evaluation), qualitative evaluations cannot move beyond responsiveness to become proactive players in the social policy-making arena.

From Whence Came Qualitative Evaluations

In the mid-1960s, evaluators were urged to use one preferred set of methodological principles and

procedures—those of the experimental model—to assess the extent to which programs had attained their goals. In keeping with the tenets of experimental science, evaluators of this era adopted stances of objectivity and believed that the results of their work would anchor social planning and policy making in a politically neutral and scientific rationality. (Greene & McClintock, 1991, p. 13)

This portrait of the early days of contemporary program evaluation is distinctive for its narrow vision and naive arrogance. What has happened in the last quarter century to so dramatically transform the theory and, to a lesser but still substantial degree, the practice of social program evaluation? Developments in two aspects of evaluation methodology stand out as major forces for change: (a) evaluation's contextual, political aspect, and (b) evaluation's philosophical, methodological aspect (Greene & McClintock, 1991). The intertwined process of reciprocal change and influence in these two aspects tells much of the story of contemporary evaluation's evolution.²

Taking Off the White Lab Coat

The failures of experimental program evaluations to contribute to the enthusiasm and innovation of the Great Society era are legendary. They include recognition of the lack of fit between the requirements of the experimental model and the exigencies of social program contexts. For example, there were serious questions raised about the ethics of denying a purportedly beneficial program to some people in order to fulfill the randomization requirement of experimental design. Moreover, as prescribed by the experimental model, early program evaluators distanced themselves from the political dimensions of their work, intentionally seeking the objective stance of "politically neutral and scientific rationality." Carol Weiss (1970, 1972, 1977) most influentially critiqued this distanced stance, arguing that it substantially underlay the marginal potency of evaluations of this era. Weiss maintained that social policy and program decision making were not rational processes to which data-based enterprises such as program evaluation could contribute the definitive piece of information. Rather, "the politics of program survival and the politics of higher policymaking accord evaluative evidence relatively minor weight in the decisional calculus" (Weiss, 1987, p. 62). In short, neither distanced objectivity nor neutral rationality was going to earn program evaluators a seat at the decision-making table.

There were other influential disjunctures between the framework of experimental science and

the contexts of program evaluation. As early as 1963, Lee Cronbach questioned the dominant focus of evaluative efforts on the stated goals and objectives of social programs, arguing that evaluation could more usefully contribute to program improvement through a focus on program planning and implementation. Over the years, Cronbach has continued to argue for a pragmatic, contextually useful role for program evaluation (see, e.g., Cronbach, 1982; Cronbach & Associates, 1980), in contrast to a scientific or theory-oriented role. As a highly respected and influential theorist, Cronbach provided arguments that were important in easing the later entry and acceptance of qualitative evaluation approaches. Cronbach's views are tellingly and engagingly represented in his now-famous debate with Donald Campbell over the relative importance in evaluation studies of external validity and contextual meaningfulness (championed by Cronbach) versus internal validity and causal claims (promoted by Campbell) (for a summary of this debate, see Mark, 1986).

Michael Scriven (1967) has also challenged the goal orientation of experimental evaluation, arguing not against a focus on program outcomes or effects, but rather against an exemption of stated program goals and objectives from evaluative scrutiny. Social programs, that is, should be evaluated according to the merit and worth of their actual effects, independent of their intended effects. Adopting this explicitly value-oriented framework for evaluation, Scriven argues, renders existing social policy and program goals themselves contestable. *In what ways does this program effectively meet an important need among the designated beneficiaries?* is a Scrivenesque evaluation question, later popularized, in theory though far less in practice, in his goal-free approach to program evaluation (Scriven, 1973).

Additionally, Ernest House (1976, 1977, 1980) has contributed a distinctive and important voice to the argument that the white-coated experimental scientist is not an appropriate role model for program evaluators. House's views blend different strands of logic and argumentation, including the fundamental grounding of evaluation work in political considerations of social distributive justice, and developments in the philosophy of science (outlined in the next section). House (1980) argues that experimental evaluation is a model for program evaluation that focuses on the truth aspect of valid conclusions of the credibility and normativity of the program (p. 251). Yet, by essence of his function, "the evaluator is engaged in a political activity. His [or her] work directly affects the distribution of basic goods in the society" (pp. 254-255). Further, as an intrinsic activity, evaluation "is intimately connected with the distribution of basic goods in the society" (p. 255).

evaluation should not only be true; it should also be just" (p. 121).

Challenging (Especially Cartesian) Foundations

A significant force for change in the form and function of program evaluation, then, were contextual challenges to the meaningfulness of experimental logic for evaluation—challenges that arose largely from within the evaluation community. Interwoven with these challenges were major fractures in the philosophical justification for experimental inquiry that permeated the evaluation community from the outside domains of the philosophy of science. In particular, the Cartesian foundationalism of positivistic science—and the concomitant premiums placed on objectivity, the proper methods, detached neutrality, and grand theory—were dethroned. Many philosophers of science came to agree that there is no place or time outside the observer from which he or she can objectively view and judge the validity of knowledge claims. Rather, all observations are imbued with the historical, theoretical, and value predispositions of the observer. Hence knowledge claims are not separable from, but rather interlocked with values; are not universal, but rather time and place bound; are not certain, but rather probabilistic and contestable (see Bernstein, 1983, for an outstanding example of these philosophical projects).

Endeavoring to be participants and not just bystanders in this Kuhnian challenge to *normal science*, program evaluators read philosophy, educated themselves about long-standing philosophical issues such as the fact/value distinction, and argued with each other at conferences and in other public forums about the intrinsic sensibility of related philosophical developments and about their relevance to the essentially practical work of evaluation. It was in this context that interpretivist philosophies and qualitative methods entered evaluation discourse. Evaluation methodologies rooted in interpretivist philosophies and incorporating qualitative methods were developed as alternatives to a rejected positivist philosophy and experimentalist methodology (Guba & Lincoln, 1981; Stake, 1980; Parlett & Hamilton, 1976; Stake, 1978). In fact, within the evaluation and other applied social science communities, much of this discourse was familiarly named the *qualitative-quantitative paradigm debate*. Although qualitative evaluations were initially contested on practical and methodological grounds, they eventually evolved to a detente (Cook & Smith, 1979; Smith & Heshusius, 1986), signifying the important acceptance of these alternative evaluation methodologies, at least among

many evaluation theorists and methodologists.³ Coming both from long-standing inquiry traditions such as ethnography, symbolic interactionism, and phenomenology and from the more recent critiques of established social science, the arguments favoring the legitimacy and potential usefulness of qualitative approaches to applied social inquiry quite simply overwhelmed, or in some cases co-opted (Gage 1989), the opposition.

Accepting Diversity in Approaches to Evaluation

It was actually the combined force of the political-contextual and the methodological-philosophical arguments that catalyzed the development and later acceptance of a diverse range of alternative approaches to program evaluation, including practical, decision-oriented approaches and approaches framed around qualitative methodologies. This diversity is now being extended, with increasing calls to recenter social program evaluation around normative concerns (Schwandt, 1989; Sirotnik, 1990), as addressed in ensuing sections of this chapter.

The current accepted legitimacy of diverse evaluation approaches is well illustrated by the *Standards for Evaluation of Educational Programs, Projects, and Materials*, originally developed in 1981 by the Joint Committee on Standards for Educational Evaluation, representing 12 professional associations, and in the process of being updated. Reflecting the importance of contextual sensitivity and methodological diversity in the evaluation field, the 30 standards developed are clustered within four critical attributes of program evaluation: utility, feasibility, propriety, and accuracy. Moreover, the standards "do not exclusively endorse any one approach to evaluation. Instead, the Joint Committee has written standards that encourage the sound use of a variety of evaluation methods. . . . Usually, it is desirable to employ multiple methods, qualitative as well as quantitative" (Stufflebeam, 1991, p. 257). In other words, "merit lies not in form of inquiry, but in relevance of information" (Cronbach & Associates, 1980, p. 7). Although praised for their openness to alternative methods, the standards have also been criticized for their decision-oriented bias, for their reliance to some degree on the conceptual vocabulary of conventional science, and hence for their only partial embracement of alternative evaluation methodologies (e.g., Linn, 1981).

With these historical notes as important context, the next two sections of this chapter more closely examine qualitative approaches to evaluation, first in their philosophical form and then in their practical form.

The Logic of Justification for Qualitative Evaluations

Just as there is no one form of qualitative evaluation practice, there is no single philosophical logic of justification universally embraced by qualitative evaluators. Yet, there is a dominant set of philosophical tenets and stances guiding qualitative evaluation fieldwork, one with both a historical legacy and a strong contemporary presence in other domains of applied social science. This philosophical inquiry framework—variously called qualitative, ethnographic, and naturalistic—is most aptly called interpretivist (Smith, 1989). This label directly connotes one of its central premises, namely, that “in the world of human experience, there is only interpretation” (Denzin, 1989, p. 8). The dominance of an interpretivist logic of justification for qualitative evaluations can be traced primarily to the highly influential work of Yvonna Lincoln and Egon Guba (1985; Guba & Lincoln, 1981, 1989). These authors have been leading advocates of alternative paradigm inquiry, particularly within the evaluation field. Consistently grounded in detailed explications of philosophical debates and developments, their work has been influential in making the philosophical premises of inquiry both visible and accessible to many in the evaluation community. Although Guba and Lincoln’s own philosophical thinking and naming have evolved over the years, core elements of their philosophical worldview have remained essentially the same. Drawing on their work and that of others, notably John Smith (1989), I will now provide an overview of the interpretivist paradigm, as generally understood and utilized by qualitative program evaluators.

The Interpretivist Paradigm in Evaluation

At root, interpretivism is about *contextualized meaning*. Interpretivist logic rejects the primacy of scientific realism, in either its traditional or more contemporary forms (House, 1991), along with its accompanying correspondence theory of truth. Rather, in interpretivism, social reality is viewed as significantly socially constructed, “based on a constant process of interpretation and reinterpretation of the intentional, meaningful behavior of people—including researchers” (Smith, 1989, p. 85), and “truth is ultimately a matter of socially and historically conditioned agreement” (p. 73). “‘Reality’ resides neither with an objective external world nor with the subjective mind of the knower, but within dynamic transactions between the two” (Barone, 1992b, p. 31). Social inquiry therefore is mind dependent; inquiry descriptions

and interpretations are themselves constructions and (re)interpretations; and there can be no separation of the investigator from the investigated (Smith, 1989, chap. 4). Interpretivist inquiry is unabashedly and unapologetically subjectivist. It is also dialectic, for the process of meaning construction transforms the constructors.

Moreover, what is important to know, what constitutes an appropriate and legitimate focus for social inquiry, is the phenomenological meaningfulness of lived experience—people’s interpretations and sense makings of their experiences in a given context. As Smith (1989) notes, this process is inevitably hermeneutical because “investigators, like everyone else, are part of the circle of interpretation” (p. 136). So understanding meaning as the goal of interpretivist inquiry “is not a matter of manipulation and control, particularly via method, but rather it is a question of openness and dialogue” (p. 137).

Meanings thus understood, or knowledge claims in interpretivist inquiry, take the form of working hypotheses or contextualized, temporary knowledge. Interpretivist knowledge claims are contestable precisely because they are contextualized and multiplistic, and also because they represent an intertwining of facts and values. There are “no facts without values, and different values can actually lead to different facts” (Smith, 1989, p. 111). In this respect, interpretivism is value laden while it is simultaneously value relative or equally malleable by inquirers with quite different value stances (Greene, 1990, 1992). That is, interpretivist practice intentionally reveals the value dimensions of lived experience (because there are no facts without values), but the dimensionality revealed is importantly connected to, even constitutive of, the value orientations and stances of the inquirer (because different values can lead to different facts). Interpretivism as a philosophical logic of justification for inquiry acknowledges, even celebrates, the permeation of values throughout the inquiry process and results, but does not advocate or prescribe any one particular set of values for social inquiry. These are thus brought by the inquirer and, in this way, the values promoted by interpretivist inquiry practice are inherently varied and diverse.

Methodologically, interpretivism is most consonant with natural settings, with the human inquirer as the primary gatherer and interpreter of meaning, with qualitative methods, with emergent and expansionist inquiry designs, and with hermeneutic understanding, in contrast to interventionist prediction and control, as the overall goal of inquiry (Guba & Lincoln, 1989; Lincoln, 1990).

Yet, perhaps more significantly, the interpretive logic of justification represents a decentering of inquiry theory and practice from questions of

method (Greene, 1992). In the field of program evaluation, which has long been characterized as, if not faulted for being, method driven, this is a substantial change. Like other interpretive inquirers, interpretivist evaluators reject the conventional stance that proper methods can insulate against bias and thereby ensure objectivity and truth. Yet, unlike some of their more radical peers, who disclaim the existence of any privileged methods that will enhance the acceptability of an inquirer's interpretations (Smith, 1989, p. 160), most interpretivist evaluators seek some procedural guidelines and support for their work. In particular, interpretivist evaluators seek to authenticate their interpretations as empirically based representations of program experiences and meanings, rather than as biased inquirer opinion. Time-honored procedures such as triangulation and negative case analysis (Denzin, 1978) and newer procedures such as member checks, peer debriefers, and audits (Lincoln & Guba, 1985) are all utilized by interpretivist evaluators to enhance the credibility of their inferences. Evaluators sense particular pressure to invoke such procedures because the contexts of program evaluation continue to demand assurances of methodological quality and data integrity in evaluative work. This work can make no contributions to social policy and program decision making unless it is perceived as credible and trustworthy.

So, although methods may not occupy center stage in interpretive evaluation approaches, questions of procedure in these approaches remain. And these questions are problematic because "to argue that certain procedures are required would simply pose a contradiction—the attempt to provide a methodological foundation for knowledge based on nonfoundational assumptions" (Smith, 1989, p. 159). On the one hand, interpretivist evaluators need methodological quality assurances for their audiences. On the other hand, the very idea of prescriptions for quality or any other methodological concern is philosophically inconsistent with the basic tenets of interpretivism. In response to this dilemma, interpretivist evaluators have generally accepted Smith's (1989, 1990) recasting of methodological concerns as choices, procedural guidelines as heuristics, and quality criteria as ever-evolving, open-ended lists.

Paradigms and Practice

Of course, not all qualitative evaluators find Smith's perspective sensible. Patton (1990), for example, offers a highly interpretivist frame for his qualitative evaluation approach in the form of "strategic themes," and simultaneously—without a single pang of philosophical conscience—promotes conventional measurement validity and re-

liability as key quality dimensions of qualitative data (p. 461). This is because Patton (1988, 1990) eschews the idea that inquiry paradigms frame or delimit methodological choices:

Rather than believing that one must choose to align with one paradigm or another, I advocate a paradigm of choices. A paradigm of choices rejects methodological orthodoxy in favor of *methodological appropriateness* as the primary criterion for judging methodological quality. The issue then becomes . . . whether one has made sensible methods decisions given the purpose of the inquiry, the questions being investigated, and the resources available. (Patton, 1990, pp. 38-39)

For Patton, the selection, design, and implementation of evaluation methods should be flexibly based on practical need and situational responsiveness, rather than on the consonance of a set of methods with any particular philosophical paradigm. And so, "objectivist" and "subjectivist" methods can be used together unproblematically. This *practical pragmatic* stand is strongly supported by other applied social inquirers (e.g., Bryman, 1988; Firestone, 1990; Pitman & Maxwell, 1992), as well as by arguments from a position of *philosophical pragmatism* (e.g., Howe, 1988).

Although clearly supporting multiplistic mixes of qualitative and quantitative methods at the methodological level, Guba and Lincoln strongly contest the mixing of inquiry approaches at the paradigm level. They argue, for example, that one cannot simultaneously adhere to the objectivist detachment of conventional science and the subjectivist involvement of interpretivism. There are others who agree that paradigms are irreconcilable, yet still seek not accommodation but dialectically enhanced inquiry benefits through a pluralistic acceptance of multiple ways of knowing. To illustrate, Salomon (1991) maintains that social issues are vastly complex and thus require both an "analytic" and a "systemic" approach to inquiry, used in a complementary fashion across studies toward more complete understanding. This essential tension between philosophical paradigms and practice is likely to remain contested. It matters to evaluators because it has important effects on how we envision and do our work, as elaborated in the next section.

The Practice of Qualitative Evaluations

There are several critical dimensions of qualitative evaluation practice. Most qualitative evaluators (a) use case studies to frame their work and

hence emphasize context, but not generalizability, as an essential element of meaning; (b) rely heavily but not exclusively on qualitative methods for meaning construction; (c) acknowledge if not celebrate the influential presence of their own selves in the inquiry process; and (d) seek in their work primarily to augment practical program understanding. The ensuing discussion will elaborate these dimensions, as well as connect them to prominent qualitative evaluation theories. In this way, salient features of each theory are highlighted, allocating more comprehensive portrayals of the theories to the works referenced.

Cases and Contexts

Some years ago, Robert Stake (1978) began encouraging evaluators to direct their energies toward the practical program concerns of stakeholders in the immediate context, rather than toward the more abstract questions of remote decision makers. Stake argued that by responsively focusing on the priority issues of practitioners within a given program or bounded case, evaluators can construct rich experiential understandings of that case. Such understandings, in turn, not only provide powerful information for program improvement, but also constitute a basis for *naturalistic generalizations*, which are grounded in the vicarious experience and tacit knowledge of the case reader. "Naturalistic generalizations develop within a person as a product of experience. They derive from the tacit knowledge of how things are . . . they seldom take the form of predictions but lead regularly to expectations. They guide action, in fact they are inseparable from action" (Stake, 1983, p. 282). Eisner (1991) makes a similar argument regarding the thematics of evaluative connoisseurship. Thematics represent the concrete universals of the case evaluated, the lessons learned, the moral of the story, and, as such, are of likely interest to others outside the case evaluated.

However, developing generalizable knowledge—even in such nonpropositional forms—is not the primary justification for embracing a case study framework for qualitative evaluations. Quite the contrary—the case for case studies in qualitative evaluations rests on a confluence of their responsive political-value stance and their underlying interpretivist assumptions. *Responsive* evaluation, first championed by Stake (1975; see also Stake, 1991), seeks expressly to uncover and then address the concerns of program stakeholders in the setting being evaluated toward the improvement of practice in that setting. From this responsive perspective, program improvement is more likely if local rather than remote concerns are addressed in the evaluation and if local rather than remote

values are explicated and used to make program judgments. Although not all qualitative evaluators maintain an exclusive focus on program practitioners, most remain within evaluation's responsive tradition. In large part, this is because this tradition is philosophically buttressed by interpretivism's view of knowledge as contextualized meaning. As an essential part of meaning, context must be described and its contributions understood.

Qualitative evaluations thus characteristically take the form of case studies, with respectful attention to context, and rarely, if ever, resemble surveys, quasi-experiments, or other inquiry formats. Deciding just what constitutes a case, however, usually requires considered judgment, involving a balancing of desired results with available resources. Rarely do evaluation resources enable an in-depth assessment of all possible cases in a program setting. In an educational program evaluation, for example, considering the whole school district as a case is unlikely to be as feasible or as useful as considering schools, or grade levels within schools, or social groups within neighborhoods as possible cases. The latter are more likely to offer differentiated understandings of such key evaluation foci as peer group norms and influences.

A Preference for Qualitative Methods

All qualitative approaches to program evaluation are distinguished by their preference for qualitative methods, including open-ended interviews, on-site observation, participant observation, and document review. For many theorists and practitioners, these methods offer the greatest consonance with the interpretivist perspective that frames and guides their work. Qualitative methods rely on the interactional, adaptive, and judgmental abilities of the human inquirer; the interpretivist challenge of understanding and interpreting meaning demands no less.

Methods choices in evaluation studies are not only influenced by philosophical assumptions and frameworks. As important, or even more important, methods choices must match the information needs of the identified evaluation audiences. For this reason, surveys, client record analysis, and other quantitative methods are commonly incorporated into interpretivist evaluations. Representing an extreme position on this matter, the influential evaluation theorist Michael Patton (1990) advocates qualitative methods when they represent the best match to the intended evaluation user's information needs, rather than because they are consonant with interpretivism. That is, Patton contends that methods choices should devolve not primarily from some abstract philosophical paradigm, but rather substantially from the concrete information needs of identified evaluation users.

When these information needs comprise multiple perspectives, contextualized meanings, or the experience of program participation, for example, then qualitative methods should be employed. This aphoristical stance on the justification for methods choices sets Patton somewhat apart from most other qualitative evaluation theorists, including Stake, Guba and Lincoln, and Eisner.

The Acknowledged Self in Inquiry

Perception of the world is perception influenced by skill, point of view, focus, language, and framework. The eye, after all, is not only a part of the brain, it is a part of tradition. . . . [So what] we know is a function of a *transaction* between the qualities of the world we cannot know in their pure, nonmediated form, and the frames of reference, personal skills, and individual histories we bring to them. . . . [Knowledge or] experience thus conceived is a form of human achievement; it is not simply had, it is made. (Eisner, 1992, pp. 11-13)

Evaluators are subjective partners with stakeholders in the literal creation of evaluation data. (Guba & Lincoln, 1989, p. 110)

Although Eisner and Guba and Lincoln differ in significant aspects of their theories, they share the premise that human knowledge is literally constructed during inquiry and hence is inevitably entwined with the perceptual frames, histories, and values of the inquirer. Qualitative evaluation à la Eisner or Guba and Lincoln is unabashedly subjective, unapologetically imbued with the individual perspectives and frames of the inquirer. No apologies are offered here, for two main reasons. First, along with many others in and outside of interpretivism, these theorists maintain that objectivity—understood as distanced detachment and neutrality intended to guard against bias and thereby to ensure the attainment of truth—is not possible and therefore should be rejected as a regulative ideal for social inquiry. Second, from an interpretivist perspective, it is precisely the individual qualities of the human inquirer that are valued as indispensable to meaning construction. In fact, Eisner's evaluation theory directly calls upon the substantive expertise of the individual connoisseur or expert. With a conjoint grounding in the arts, this theory highlights the enlightened eye and the seasoned judgment of the inquirer, along with his or her expertise in representation or in making public what has been seen. Eisner's evaluation connoisseur is an expert in the program to be assessed. Relying heavily on qualitative methods, the connoisseur collects information and then uses his or her expert frames and insights to integrate, interpret, and judge.

Thus the self of the qualitative evaluator is acknowledged to be present in the inquiry, a presence that permeates all methodological decisions and penetrates the very fabric of meaning constructed. Yet, just how much influence is acceptable and just how such influence can be monitored or detected, as needed, are issues of little consensus within the qualitative evaluation community. Eisner would contend, for example, that as the connoisseur, the qualitative evaluator carries authority that is supreme, and his or her influence in all aspects of the inquiry is expressly valued. In contrast with his view of the evaluator as a technical consultant, Patton would allocate more substantive authority to evaluation users. And he would rely on the proper application of qualitative methods to minimize the evaluator's presence in the program experiences and meanings investigated and understood. Given their vision of the evaluator as a negotiator, Guba and Lincoln would take a middle position, arguing that the meanings created and the interpretation and use of those meanings are responsibilities shared among the evaluator and program stakeholders. This range of views on the desired presence of self in qualitative evaluation is reflected in varied views on its primary purpose, as discussed next.

The Envisioned Inquiry in Society

Some vision of purpose is, at root, what guides all evaluation practice. In the responsive tradition in evaluation, inquirers have sought to augment local program understanding with the hope of moving toward program improvement. And the primary audiences for responsive evaluation have been program practitioners and participants. This vision of evaluation as the generation of local contextualized insights, as the reflective sharing of new program perspectives, as the telling of diverse program stories, is highly congruent with an interpretivist philosophy and a qualitative methodology. For example, participant observation is ideally suited for constructing the emic meaning of program participation for varied participants in that particular context.

For some within the evaluation community, however, understanding emic meaning and relating diverse program stories is not enough. Rather, as social inquirers in the public domain, we are morally and ethically compelled to assume greater responsibility for and a more active role in the social policy arena. Because our work can affect who gets what (House, 1980), we must be actively engaged with the consequences of our work. Yet, because of its inherent value relativity, interpretivism—while permitting alternative visions

of evaluation purpose—does not in and of itself provide sufficient guidance or warrant for any particular alternative vision. Such guidance and warrant must come from somewhere outside the interpretivist logic of justification.

One important, long-standing exemplar of such politically engaged evaluation is the democratic evaluation tradition promoted by MacDonald (1976), Simons (1987), McTaggart (1990), and others. This evaluation model most centrally seeks to balance the public's right to know with the individual's right to privacy and to be discrete. The balance is sought via prescribed procedures for accessing, reviewing, negotiating, and releasing evaluation data. The concept of democratic evaluation is derived from the tradition of liberal democracy, and is thus politically and morally acceptable to existing power holders in democratic societies. At the same time, democratic evaluation seeks within its own boundaries to forge power-equalizing interactions and to establish a flow of information that is independent of hierarchical interests. Hence, in a democratic evaluation context, all relevant perspectives can be represented, information can be fairly and equitably exchanged, and open deliberation can be encouraged (Simons, 1987).

Guba and Lincoln's (1989) more recent fourth-generation evaluation approach represents a purposeful blending of an essentially interpretivist philosophy with an outside warrant for social action into what they call a *constructivist* framework for evaluation. Moving beyond the telling of stories, constructivism requires that evaluation catalyze social action. Yet, in consonance with interpretivism, the specific contours and facets of that action are not prescribed but rather emerge from the setting. In this way, constructivism differs from the more prescriptive empowerment, equity, and social justice agendas of critical, feminist, and other normative inquiry approaches. With this infusion of an outside warrant, fourth-generation evaluation seeks not so much program understanding as social change-oriented action, and the fourth-generation evaluator's role is not so much one of describer and consultant as one of negotiator and social change catalyst. More significantly, Guba and Lincoln's fourth-generation departures from the responsive tradition in evaluation well illustrate the critical vulnerabilities of qualitative evaluation today. These vulnerabilities are primarily in the areas of purpose and audience, in the envisioned role of inquiry in society. Interpretivism envisions no particular role, although an increasing clamor of contemporary voices is insisting otherwise. A sampling of these voices and what they portend for qualitative evaluation in the years ahead conclude this chapter.

Continuing Challenges

Given the inescapable incursion of values into human activity, Freire's . . . dictum that there can be no neutral education is extended to practices of social inquiry. The inescapable political content of theories and methodologies becomes increasingly apparent. (Lather, 1992, p. 4)

Social justice is among the most important values we should hope to secure in evaluation studies. . . . Public evaluation should be an institution for democratizing public decision making. . . . As a social practice, evaluation entails an inescapable ethic of public responsibility . . . [serving] the interests of the larger society and of various groups within society, particularly those most affected by the program under review. (House, 1990, pp. 23-24)

Of pivotal significance to the dizzying pluralism of social inquiry in the present era is the recognition that values permeate all observations, and hence all methods that are used to gather the observations, and hence all methodologies that frame and guide implementation of the methods. A critical question of the era is, thus, What values or whose? And in the public contexts of social program evaluation, the question becomes, What societal values or what visions of community constitute warranted frames for evaluation methodology?

Ernest House offers a vision rooted in conceptions of social justice (see also Sirotnik, 1990). Thomas Schwandt (1989, 1991) argues for a morally engaged evaluation practice that "aims at achieving insight and awareness into what it means to live a human life" (1991, p. 70). Schwandt's ideas incorporate recent challenges to the liberalist political tradition and concomitant arguments to replace our failed representative democracy with a genuinely participatory one (Barber, 1984) or with a communitarian ethic rooted in human interdependence and solidarity (Sullivan, 1986). An additional emancipatory vision for applied social inquiry is broadly represented by critical social scientists (e.g., Fay, 1987) and, more tellingly, by practicing participatory evaluators (Brunner & Guzman, 1989; Whitmore, 1990). In this vision, evaluation is viewed primarily as a process for promoting empowerment and requisite structural change. All of these visions offer an explicit political agenda for program evaluation, thereby not only recognizing the presence of values in inquiry, but specifically promoting one particular normative frame.

Of key relevance to the present discussion are questions of the intersection between this "inescapable" normative turn in social inquiry and

qualitative approaches to program evaluation. In response, I would argue that qualitative approaches are highly compatible with and hence have much to offer more openly ideological approaches to program evaluation at the level of method or technique. Qualitative methods, for example, can effectively give voice to the normally silenced and can poignantly illuminate what is typically masked. Qualitative methods and approaches are already being employed in such ideologically oriented inquiry frameworks as critical ethnography (Anderson, 1989), narrative inquiry (Barone, 1992a; Bruner, 1986), and feminist social science (Lather, 1992). To illustrate, Fine (Chapter 4, this volume), notes, "More interestingly, qualitative researchers have begun to interrupt Othering by forcing subjugated voices in context to the front of our texts and by exploiting privileged voices to scrutinize the technologies of Othering." And Lather (1992) states:

Feminist methodologies and epistemologies, [Harding] suggests, require new feminist uses of these familiar research methods . . . [of] listening to informants, observing behavior, [and] examining historical records. . . . Studying women from the perspective of their own experiences so that they/we can better understand our situations in the world is research designed for women instead of simply research *about* women. (p. 6)

At the broader level of methodology or paradigm, however, the interpretivist framework does not provide sufficient warrant or guidance for any given normative agenda. Interpretivism justifies values in inquiry, but does not justify any particular ones. As participants in the social policy arena, program evaluators are increasingly being called upon to get involved, to be a part of the action, to become public scientists. With their acknowledgment of values, qualitative approaches can help evaluators illuminate alternative paths or courses of action. Such approaches can be molded to fit varied and emerging inquiry shapes, from technical reports to dramatic dialogue. And they can adaptively respond to varied and evolving inquiry functions, including shifting social action agendas. For these reasons, qualitative approaches are likely to continue to be a significant and useful alternative in the methodological repertoire of program evaluators. Yet, also for these reasons, qualitative evaluations as a genre are destined to remain within evaluation's responsive tradition—beautifully responsive but, in being so, unable to assume a more proactive role in the social policy sphere. And so, because the evaluator as public scientist must be proactive, must him- or herself become an active and accountable player in the policy arena, qualitative evaluations will not be enough.

Notes

1. Evaluations are also conducted on objects (product evaluation) and on people (personnel evaluation), as well as on programs that few would consider social (e.g., executive professional development via outdoor experiences). Although Scriven (1991) and others would argue that the logic of evaluation is the same across different forms and objects of evaluation, I believe they constitute radically different tasks requiring qualitatively and politically different responses. This chapter, therefore, is restricted to social program evaluation, predominantly in the public domain.

2. Most major evaluation texts also have chapters on evaluation's history. See, for example, Guba and Lincoln (1981, 1989); Madaus, Scriven, and Stufflebeam (1983); Patton (1986); Rossi and Freeman (1985); and Shadish, Cook, and Leviton (1991).

3. That this debate runs deep and long is attested to by the recent sequence of Presidential Addresses by the 1990 and 1991 presidents of the American Evaluation Association, an exchange of quantitative and qualitative views that rekindled emotional layers of this debate (Lincoln, 1991; Sechrest, 1992). Moreover, as noted previously, important sectors of the evaluation community, including many practitioners and audiences, missed the qualitative-quantitative debate entirely. Admittedly, this debate can certainly be viewed as a rarefied intellectual exchange of no relevance to daily life. Nonetheless, there remains a gap between evaluation theory and practice, with much of the latter based on discarded epistemologies and paradigms. This gap remains troublesome to evaluators concerned about both defensible theory and meaningful practice.

References

- Anderson, G. L. (1989). Critical ethnography in education: Origins, current status, and new directions. *Review of Educational Research, 59*, 249-270.
- Atkinson, P., Delamont, S., & Hammersley, M. (1988). Qualitative research traditions: A British response to Jacob. *Review of Educational Research, 58*, 231-250.
- Barber, B. (1984). *Strong democracy: Participatory politics for a new age*. Berkeley: University of California Press.
- Barone, T. E. (1992a). Beyond theory and method: A case of critical storytelling. *Theory Into Practice, 31*(2), 142-146.
- Barone, T. E. (1992b). On the demise of subjectivity in educational inquiry. *Curriculum Inquiry, 22*, 25-38.
- Bernstein, R. J. (1983). *Beyond objectivism and relativism*. Philadelphia: University of Pennsylvania Press.
- Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

- Brunner, I., & Guzman, A. (1989). Participatory evaluation: A tool to assess projects and empower people. In R. F. Conner & M. Hendricks (Eds.), *International innovations in evaluation methodology* (pp. 9-18). San Francisco: Jossey-Bass.
- Bryman, A. (1988). *Quantity and quality in social research*. London: Unwin Hyman.
- Chen, H., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review*, 7, 283-302.
- Cicirelli, V. G., & Associates. (1969). *The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development* (Report to the Office of Economic Opportunity). Athens, OH: Ohio University/Westinghouse Learning Corporation.
- Cook, T. D. (1985). Postpositivist critical multiplism. In L. Shotland & M. M. Mark (Eds.), *Social science and social policy* (pp. 21-62). Beverly Hills, CA: Sage.
- Cook, T. D., & Reichardt, C. S. (Eds.). (1979). *Qualitative and quantitative methods in evaluation research*. Beverly Hills, CA: Sage.
- Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, 64, 672-683.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Cronbach, L. J., & Associates. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.
- Denzin, N. K. (1978). *The research act: An introduction to sociological methods*. New York: McGraw-Hill.
- Denzin, N. K. (1989). *Interpretive interactionism*. Newbury Park, CA: Sage.
- Eisner, E. W. (1976). Educational connoisseurship and criticism: Their forms and functions in educational evaluation. *Journal of Aesthetic Education*, 10, 135-150.
- Eisner, E. W. (1991). Taking a second look: Educational connoisseurship revisited. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (pp. 169-187). Chicago: University of Chicago Press.
- Eisner, E. W. (1992). Objectivity in educational research. *Curriculum Inquiry*, 22, 9-15.
- Fay, B. (1987). *Critical social science*. Ithaca, NY: Cornell University Press.
- Firestone, W. A. (1990). Accommodation: Toward a paradigm-praxis dialectic. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 105-124). Newbury Park, CA: Sage.
- Gage, N. L. (1989). The paradigm wars and their aftermath: A "historical" sketch of research on teaching since 1989. *Educational Researcher*, 18(7), 4-10.
- Greene, J. C. (1990). Three views on the nature and role of knowledge in social science. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 227-245). Newbury Park, CA: Sage.
- Greene, J. C. (1991). *Responding to evaluation's moral challenge*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Greene, J. C. (1992). The practitioner's perspective. *Curriculum Inquiry*, 22(1), 39-45.
- Greene, J. C., & McClintock, C. (1991). The evolution of evaluation methodology. *Theory Into Practice*, 30(1), 13-21.
- Guba, E. G. (Ed.). (1990). *The paradigm dialog*. Newbury Park, CA: Sage.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- House, E. R. (1976). Justice in evaluation. In G. V Glass (Ed.), *Evaluation studies review annual* (Vol. 1). Beverly Hills, CA: Sage.
- House, E. R. (1977). *The logic of evaluative argument*. Los Angeles: University of California, Center for the Study of Evaluation.
- House, E. R. (1980). *Evaluating with validity*. Beverly Hills, CA: Sage.
- House, E. R. (1990). Methodology and justice. In K. A. Sirotnik (Ed.), *Evaluation and social justice* (pp. 23-36). San Francisco: Jossey-Bass.
- House, E. R. (1991). Realism in research. *Educational Researcher*, 20(6), 2-9.
- Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational Researcher*, 17(8), 10-16.
- Jacob, E. (1987). Qualitative research traditions: A review. *Review of Educational Research*, 57, 1-50.
- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.
- Lather, P. (1992). Critical frames in educational research: Feminist and poststructural perspectives. *Theory Into Practice*, 31(2), 1-13.
- Lincoln, Y. S. (1990). The making of a constructivist: A remembrance of transformations past. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 67-87). Newbury Park, CA: Sage.
- Lincoln, Y. S. (1991). The arts and sciences of program evaluation (AEA 1991 Presidential Address). *Evaluation Practice*, 12(1), 1-7.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Linn, M. (1981). Standards for evaluating out-of-school learning. *Evaluation News*, 2(2), 171-176.
- MacDonald, B. (1976). Evaluation and the control of education. In D. A. Tawney (Ed.), *Curriculum evaluation today: Trends and implications*. London: Falmer.
- Madaus, G. F., Scriven, M., & Stufflebeam, D. L. (Eds.). (1983). *Evaluation models: Viewpoints on educa-*

- ional and human services evaluation. Boston: Kluwer-Nijhoff.
- Mark, M. M. (1986). Validity typologies and the logic and practice of quasi-experimentation. In W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis*. San Francisco: Jossey-Bass.
- McTaggart, R. (1990, April). *Dilemmas in democratic evaluation: Politics and validation*. Paper presented at the annual meeting of the American Educational Research Association, Boston.
- Parlett, M., & Hamilton, D. (1976). Evaluation as illumination: A new approach to the study of innovative programs. In G. V. Glass (Ed.), *Evaluation studies review annual* (Vol. 1). Beverly Hills, CA: Sage.
- Patton, M. Q. (1986). *Utilization-focused evaluation* (2nd ed.). Beverly Hills, CA: Sage.
- Patton, M. Q. (1987). Evaluation's political inherency: Practical implications for design and use. In D. J. Palumbo (Ed.), *The politics of program evaluation* (pp. 100-145). Newbury Park, CA: Sage.
- Patton, M. Q. (1988). Paradigms and pragmatism. In D. M. Fetterman (Ed.), *Qualitative approaches to evaluation in education: The silent scientific revolution* (pp. 116-137). New York: Praeger.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.
- Pitman, M. A., & Maxwell, J. A. (1992). Qualitative approaches to evaluation: Models and methods. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 729-770). New York: Academic Press.
- Rossi, P. H., & Freeman, H. E. (1985). *Evaluation: A systematic approach* (3rd ed.). Beverly Hills, CA: Sage.
- Rossi, P. H., & Lyall, K. C. (1978). An overview of the NIT experiment. In T. D. Cook, M. L. DelRosario, K. M. Hernigan, M. M. Mark, & W. M. K. Trochim (Eds.), *Evaluation studies review annual* (Vol. 3, pp. 412-428). Beverly Hills, CA: Sage.
- Salomon, G. (1991). Transcending the qualitative-quantitative debate: The analytic and systemic approaches to educational research. *Educational Researcher*, 20(6), 10-18.
- Schwandt, T. A. (1989). Recapturing moral discourse in evaluation. *Educational Researcher*, 18(8), 11-16, 34.
- Schwandt, T. A. (1991). Evaluation as moral critique. In C. L. Larson & H. Preskill (Eds.), *Organizations in transition: Opportunities and challenges for evaluation* (pp. 63-72). San Francisco: Jossey-Bass.
- Scriven, M. (1967). The methodology of evaluation. *AERA Monograph Series in Curriculum Evaluation*, 1, 39-83.
- Scriven, M. (1973). Goal-free evaluation. In E. R. House (Ed.), *School evaluation: The politics and process* (pp. 319-328). Berkeley, CA: McCutchan.
- 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). Beyond formative and summative evaluation. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (pp. 19-64). Chicago: University of Chicago Press.
- Sechrest, L. (1992). Roots: Back to our first generations (AEA 1991 Presidential Address). *Evaluation Practice*, 13(1), 1-7.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
- Simons, H. (1987). *Getting to know schools in a democracy: The politics and process of evaluation*. London: Falmer.
- Sirotnik, K. A. (Ed.). (1990). *Evaluation and social justice*. San Francisco: Jossey-Bass.
- Sirotnik, K. A., & Oakes, J. (1990). Evaluation as critical inquiry: School improvement as a case in point. In K. A. Sirotnik (Ed.), *Evaluation and social justice* (pp. 37-53). San Francisco: Jossey-Bass.
- Smith, J. K. (1989). *The nature of social and educational inquiry: Empiricism versus interpretation*. Norwood, NJ: Ablex.
- Smith, J. K. (1990). Alternative research paradigms and the problem of criteria. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 167-187). Newbury Park, CA: Sage.
- Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate. *Educational Researcher*, 15(1), 4-12.
- Soltis, J. F. (1990, April). *The hermeneutics/interpretive tradition and its virtues*. Paper presented at the annual meeting of the American Educational Research Association, Boston.
- Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record*, 68, 523-540.
- Stake, R. E. (1975). *Evaluating the arts in education: A responsive approach*. Columbus, OH: Merrill.
- Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher*, 7(1), 5-8.
- Stake, R. E. (1983). The case study method in social inquiry. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 279-286). Boston: Kluwer-Nijhoff.
- Stake, R. E. (1991). Retrospective on "The countenance of educational evaluation." In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (pp. 67-88). Chicago: University of Chicago Press.
- Stufflebeam, D. L. (1991). Professional standards and ethics for evaluators. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At*

- quarter century* (pp. 249-282). Chicago: University of Chicago Press.
- Sullivan, W. M. (1986). *Reconstructing public philosophy*. Berkeley: University of California Press.
- Weiss, C. H. (1970). The politicization of evaluation research. *Journal of Social Issues*, 26, 57-68.
- Weiss, C. H. (Ed.). (1972). *Evaluating action programs: Readings in social action and education*. Boston: Allyn & Bacon.
- Weiss, C. H. (Ed.). (1977). *Using social research in public policy making*. Lexington, MA: Lexington.
- Weiss, C. H. (1987). Where politics and evaluation research meet. In D. J. Palumbo (Ed.), *The politics of program evaluation* (pp. 47-70). Newbury Park, CA: Sage.
- Whitmore, E. (1990). *Focusing on the process in evaluation: It's the "how" that counts*. Paper presented at the annual meeting of the American Evaluation Association, Washington, DC.