

A World Larger than Formative and Summative

MICHAEL QUINN PATTON

ABSTRACT

Patton continues the debate by identifying three arenas of evaluation practice in which the formative/summative dichotomy appears limited: knowledge-generating evaluations aimed at conceptual rather than instrumental use; developmental evaluation; and use of evaluation processes to support interventions or empower participants. In so doing, the essence of "evaluation" is more broadly defined, and the impact of harsh criticism on the listener is demonstrated through personal example.

Patton

INTRODUCTION

The article that Scriven (1991) wrote, to which this is an invited response, is lengthy and detailed. My response will focus on only a few of the major points that concern the central issue posed to me: Whether all the possible purposes of evaluation can be captured by the single, dichotomous distinction between formative and summative evaluations. My position, as the title suggests, is that the world of evaluation has grown larger than the boundaries of formative and summative evaluation, though this classic distinction remains important and useful.

Scriven does, indeed, posit that the formative versus summative distinction captures the entire array of evaluation purposes. He says:

Formative evaluation is evaluation designed, done, and intended to support the process of improvement, and normally commissioned or done by, and delivered to, someone who can make improvements. *Summative evaluation* is the rest of evaluation: in terms of intentions, it is evaluation done for, or by, any observers or decision makers (by contrast with

Michael Quinn Patton • Union Institute, 3228 46th Avenue South, Minneapolis, MN 55406.

Evaluation Practice, Vol. 17, No. 2, 1996, pp. 131–144.
ISSN: 0886-1633

Copyright © 1996 by JAI Press Inc.
All rights of reproduction in any form reserved.

developers) who need evaluative conclusions for any reasons besides development (Scriven, 1991, p. 20).

By defining formative and summative in this way, Scriven makes the distinction exhaustive of all possibilities because, by definition, anything that is not formative is summative. It is clear from Scriven's subsequent discussion of the relationship between formative and summative, however, that the purpose of formative evaluation is primarily to get a developing program ready for summative assessment, and that formative evaluation rests in the shadow of summative. For example, he says "the formative evaluation should *at least* provide a preview of a summative evaluation" (Scriven, 1991, p. 28; emphasis in the original). Later, he reiterates this point in saying "the formative evaluation is worth nothing at all unless it at least includes a preview of good summative evaluation" (Scriven, 1991 p. 53).

In essence, this view derives from Scriven's perspective that the only real evaluation is ultimately a fundamental judgment of merit or worth at the summative level. It is my contention, in contrast, that the field of evaluation is much broader and richer than this judgment-focused definition permits. Let me suggest three examples of important evaluation use that are not encompassed in the formative and summative categories.

Knowledge Generation

An increasingly important evaluation purpose that goes beyond formative-summative evaluation is the emergent area of knowledge-generating evaluation research for *conceptual* use. Both judgment-oriented (summative) and improvement-oriented (formative) evaluations involve the *instrumental* use of results (Leviton & Hughes, 1981). Instrumental use occurs when a decision or action follows, at least in part, from the evaluation. Evaluations are seldom the sole basis for subsequent summative decisions or program improvements, but they contribute, often substantially.

Conceptual use of findings, on the other hand, contrasts with instrumental use in that no decision or action is expected; rather, it "is the use of evaluations to influence thinking about issues in a general way" (Rossi & Freeman, 1985, p. 388). The evaluation *research* findings contribute by increasing knowledge. This knowledge can be as specific as clarifying a program's model, testing theory, distinguishing types of interventions, figuring out how to measure outcomes, generating lessons learned, and/or elaborating policy options. In other cases conceptual use is more vague, with users seeking to better understand the program; the findings, then, may reduce uncertainty, offer illumination, enlighten funders and staff about what participants really experience, enhance communications, and facilitate sharing of perceptions. In early studies of utilization, such uses were overlooked or denigrated. In recent years, they have come to be more appreciated and valued (Weiss, 1990, p. 177).

We found conceptual use to be widespread in our follow-up study of federal health evaluations (Patton, 1986). As one project manager reported:

The evaluation led us to redefine some target populations and rethink the ways we connected various services. This rethinking happened over a period of months as we got a better perspective on what the findings meant. But we didn't so much change what we were doing as we changed how we thought about what we were doing. That has had big pay-offs over time. We're just a lot clearer now.

This represents an example of conceptual use that is sometimes described as "enlightenment." Carol Weiss has used this term to describe the effects of evaluation findings being dis-

seminated to the larger policy community “where they have a chance to affect the terms of debate, the language in which it is conducted, and the ideas that are considered relevant in its resolution” (Weiss, 1990, p. 176).

While Weiss has emphasized the informal ways in which evaluation findings provide, over time, a knowledge base for policy, Chen has focused on a more formal knowledge-oriented approach in what he has called “theory-driven evaluation” (Chen, 1989, 1990; Chen & Rossi, 1987). Though theory-driven evaluations can provide program models for summative judgment or ongoing improvement, the connection to social science theory also offers the potential for increasing knowledge about how effective programs work in general. For example, Shadish (1987, p. 94) has argued that the understandings gleaned from evaluations ought to contribute to “macrotheories” about “how to produce important social change.” Scheirer (1987, p. 59) has contended that evaluators ought to draw on and contribute to “implementation theory” to better understand the “what and why of program delivery” Such knowledge-generating efforts focus beyond the effectiveness of a particular program to future program designs and policy formulation in general.

As the field of evaluation has matured and a vast number of evaluations has accumulated, the opportunity has arisen to look across findings about specific programs to formulate generalizations about effectiveness. This involves synthesizing findings from different studies. [It is important to distinguish this form of synthesis evaluation, that is, synthesizing across different studies, from what Scriven (1994) calls “the final synthesis,” which refers to sorting out and weighing the findings in a single study to reach a summative judgment.] Cross-study syntheses have become an important contribution of the U.S. General Accounting Office in providing accumulated wisdom to Congress about how to formulate effective policies and programs (GAO, 1989). An example is GAO’s report (1992) on “Adolescent Drug Use Prevention: Common Features of Promising Community Programs.” An excellent and important example of synthesis evaluation is Lisbeth Schorr’s, *Within Our Reach* (1988), a study of programs aimed at breaking the cycle of poverty in which she identified “the lessons of successful programs.”

Such syntheses constitute accumulated wisdom—principles of effectiveness—that can be adapted, indeed, must be adapted, to specific programs, or even entire organizations. For example, the Ford Foundation commissioned an evaluation of its Leadership Program for Community Foundations. This study of 27 community foundations over five years led to a guide for “Building Community Capacity” (Mayer, 1994) that incorporates lessons learned and generalizable development strategies for community foundations—a distinguished and useful example of a knowledge-generating evaluation. Another example is a special evaluation issue of *Marriage and Family Review* devoted to “Exemplary Social Intervention Programs” (Guttman & Sussman, 1995).

In the philanthropic world this approach is often called “cluster evaluation” (Council on Foundations, 1993, pp. 232–251). A cluster evaluation team visits a number of different grantee projects with a similar focus (e.g., grassroots leadership development) and draws on individual grant evaluations to identify patterns across and lessons from the whole cluster (Campbell, 1994; Sanders, 1994; Worthen, 1994; Barley & Jenness, 1993; Kellogg Foundation, n.d.). The McKnight Foundation commissioned a cluster evaluation of 34 separate grants aimed at aiding families in poverty. One lesson learned was that “effective programs have developed processes and strategies for learning about the *strengths* as well as the needs of families in poverty” (Patton, et al., 1993, p. 10).

Such generalizable evaluation findings about principles of effective programming have become the knowledge base of our profession. Being knowledgeable about patterns of program effectiveness allows evaluators to provide guidance about development of new initiatives, policies, and strategies for implementation. *Such contributions constitute the conceptual use of evaluation research findings—a use beyond formative and summative evaluation for specific programs.*

Synthesis evaluations also help us generate knowledge about conducting useful evaluations. The premises of utilization-focused evaluation (Patton, 1997) originally emerged from studying twenty federal evaluations (Patton, et al., 1977). Those premises have been affirmed by Alkin, et al. (1979), in the model of evaluation use they developed by analyzing evaluations from different education districts in California, and by Wargo's "characteristics of successful program evaluations," which were identified by studying three "unusually successful evaluations of national food and nutrition programs" (Wargo, 1989, p. 71). The Council on Foundations commissioned a synthesis evaluation based on nine case studies of major foundation evaluations to learn lessons about "effective evaluating." Among the Council's 35 key lessons learned is this utilization-focused evaluation premise: "**Key 6. Make sure the people who can make the most use of the evaluation are involved as stakeholders in planning and carrying out the evaluation**" (Council on Foundations, 1993, p. 255).

Knowledge-generating evaluation research is not conducted to judge the merit or worth of individual programs, but rather to generate knowledge about program (or evaluation) effectiveness in general. Such efforts may appropriately be called "evaluation *research*" in contrast to program evaluation, but such research is ultimately evaluative. Judgments are made about what works generically, but such judgments are neither formative, in the sense in which Scriven uses the word, nor are they summative. I think it's useful to maintain the terms "formative" and "summative" for specific programs and discrete decisions, and to recognize that knowledge generation for conceptual use is an important and different purpose of evaluation research.

Developmental Evaluation

A second area in which the formative-summative distinction seems insufficient is what I have come to call "developmental evaluation" (Patton, 1994, 1997). Developmental evaluation involves ongoing work with a program for continuous improvement. Scriven's original formative-summative distinction is really premised on pilot programs and demonstration models in which a program engages in formative evaluation for a specified period, perhaps two to three years, and then the resulting effort is evaluated summatively. Many programs are ongoing, however, and do not face a summative decision. Indeed, they don't even want to reach the kind of stable model necessary for rigorous summative assessment.

For example, I have been working for over ten years with some philanthropic foundation programs in areas where they have made long-term commitments and do not want or need to make judgment about continuation. They do want to continue improving the effort over the long-term, however. A variety of evaluation issues have been addressed over these years and each effort has contributed to changes in the program. Moreover, it is the philosophy of the program staff that the program must continue to develop in response to changed conditions and new knowledge. Such changes do not mean that what was done before was ineffective; rather, it means that as the world changes, the program must change. Developmental evalua-

tion supports such ongoing development. This is very different from conducting formative evaluation in an effort to get ready for summative evaluation.

But is this really different from formative evaluation?

Scriven's original distinction between formative and summative in curriculum evaluation (1967) defined the purpose of formative evaluation as getting ready for summative evaluation. That definition is certainly narrower than what I'm calling developmental evaluation. But it is also true that, over time, the meaning of formative evaluation has been enlarged to include any evaluation whose primary purpose is program improvement. But connotatively, it seems to me, formative evaluation remains connected to summative evaluation and defines a preliminary stage of assessment, while the program and its evaluation are getting ready for the *real* thing: summative judgment. That's how I read and interpret Scriven. That's why I prefer (and need) a separate term for long-term improvement-oriented evaluation in which development is an end in itself and the evaluator plays a role beyond rendering judgments about merit or worth. To elaborate this point, I'll reproduce some of the explanation I made in originally offering the term "developmental evaluation" (Patton, 1994; Stockdill, et al., 1992).

Developmental evaluation serves development-oriented programs that have as their purpose the vague, general notion of development. The process is the outcome. They eschew clear, specific and measurable goals up-front because clarity, specificity and measurability are limiting. They've identified an issue or problem and want to explore some potential solutions or interventions, but they realize that where they end up will be different for different participants—and that participants themselves should play a major role in goal-setting. The process often involves engaging participants in setting and achieving their own goals. Program designers observe where they end up and make judgments about the implications of what has happened for future programming and re-engineering. They never expect to arrive at a steady state of programming because they're constantly tinkering as participants, conditions, learnings, and context change. They don't aspire to arrive at a model subject to summative evaluation and generalization. Rather, they aspire to continuous progress, ongoing adaptation, and rapid responsiveness. No sooner do they articulate and clarify some aspect of the process than that very awareness becomes an intervention and acts to change what they do. They don't value traditional characteristics of summative excellence such as standardization of inputs, consistency of treatment, uniformity of outcomes, and clarity of causal linkages. They assume a world of multiple causes, diversity of outcomes, inconsistency of interventions, interactive effects at every level—and they find such a world exciting and desirable. They never expect to conduct a summative evaluation because they never expect the program—or world—to hold still long enough for summative review. They expect to be forever developing and changing—and they want an evaluation approach that supports development and change.

Moreover, they don't conceive of development and change as necessarily improvements. Formative evaluation carries a bias about making something better, rather than just making it different. From a developmental perspective, you do something different because something has changed—your understanding, the characteristics of participants, technology, or the world. Those changes are dictated by your current perceptions, but the commitment to change doesn't carry a judgment that what was done before was inadequate or less effective. Change is not necessarily progress. Change is adaptation. As one design team member said:

We did the best we knew how with what we knew and the resources we had. Now we're at a different place in our development—doing and thinking different things. That's development. That's change. But it's not necessarily improvement.

The developmental perspective, as I experience it, feels quite different from the traditional logic of programming in which goals are predetermined and plans are carefully made for achieving those goals. Developmental programming calls for developmental evaluation in which the evaluator becomes part of a design team helping to monitor what's happening, both processes and outcomes, in an evolving, rapidly changing environment of constant feedback and change. These relationships can go on for years and, in many cases, never involve formal, written reports.

Developmental evaluation isn't a model. It's a relationship founded on a shared purpose: development. What I bring to the design team is evaluation logic, knowledge about effective programming based on evaluation wisdom, and some methods expertise to help set up monitoring and feedback systems. I become part of the team, not apart from the team. I participate in decision-making about the program and facilitate discussion about how to evaluate whatever happens. All team members render evaluation judgments together and decide how to apply the implications of results for the next stage of development.

One reaction I've had from colleagues, including Scriven, is that these examples aren't "evaluations" at all, but rather organizational development efforts. But what I'm reporting is that, in my experiences (see Patton, 1994 and 1997 for examples), my participation, identity and role are considered evaluative by those with whom I'm engaged (and by whom I'm paid). To be sure, there is no pretense of external independence. My role varies from being evaluation facilitator to full team member. This may be a different role for an "evaluator," but I include "developmental evaluation" among the things we evaluators can do, and need a label for, because an increasing number of evaluators are engaged in this kind of activity and need a way to talk about it. It goes well beyond the narrower meaning of formative evaluation. Thus, I offer this *definition of developmental evaluation*:

Developmental evaluation involves evaluation processes, including asking evaluative questions and applying evaluation logic, to support program, project, product, staff and/or organizational development. The evaluator is part of a team whose members collaborate to conceptualize, design and test new approaches in a long-term, on-going process of continuous improvement, adaptation and intentional change. The evaluator's primary function in the team is to elucidate team discussions with evaluative questions, data and logic, and to facilitate data-based decision-making in the developmental process.

Experienced evaluators know a lot about patterns of effectiveness and we'll know more over time because of the knowledge-generating kinds of evaluations I discussed earlier. That knowledge resides in the profession and makes us valuable partners in the design process. Crossing that line, however, does reduce independence of judgment. The costs and benefits of such a role change must be openly acknowledged and carefully assessed.

Using the Evaluation Process

The third arena in which I find the formative-summative distinction insufficient is in describing and thinking about uses of evaluation processes rather than evaluation findings, a notion that grows out of developmental evaluation, but goes beyond it (Patton, 1997). As I suggested earlier, the formative versus summative distinction actually concentrates on how *findings* are used. One of the things we have learned in recent years is that much of the impact of evaluation comes from the process of engaging in the evaluation, not from the findings. Working with a program to understand the logic of evaluation, to engage in goals clarification,

to develop the program's model, and otherwise engaging staff, through evaluation, to think about and become more effective in their program efforts, can be a purpose of evaluation.

The logic and principles of evaluation can also be useful in negotiations between parties with different perspectives. For example, a major foundation was interested in funding an effort to make schools more racially equitable. The school district expressed great interest in such funding, but resisted committing to explicit school changes that might undermine building-level autonomy or intrude into personnel evaluations of principals. Over a period of several months the funder and school officials negotiated the project. The negotiations centered on expected evaluation outcomes. The funder and school district eventually agreed to focus the project and evaluation on community-based, school-specific action plans, activities, and changes rather than a standardized and prescribed set of district-determined mandates. Case studies were chosen as the most appropriate evaluation method rather than standardized instruments for measuring school climate. The design of the entire project was changed and made more focused as a result of these negotiations. Applying the logic of evaluation had a major impact on the project's design without any data collection, findings, or a report. Everyone came out of the negotiations clear about what was to happen in the project and how it would be evaluated.

From strategic planning in large organizations to internal memoranda in the smallest agencies, the evaluation principles of clarity and specificity can enhance communications. Aubel (1993, p. 13) has advocated participatory approaches to evaluation for "improved communication between staff at different levels of program implementation." Evaluators' skills in facilitating the development of explicit criteria can, in and of itself, be useful to participants in a planning or evaluation exercise.

A different use of evaluation to enhance communications and mutual understanding involves designing the evaluation to "give voice" to the disenfranchised, underprivileged, poor, and others outside the mainstream (Weiss & Greene, 1992, p. 145). In the evaluation of a diversity project in the Saint Paul Schools, a major part of the design included capturing and reporting the experiences of people of color. Providing a way for African-American, Native American, Chicano-Latino, and Hispanic parents to tell their stories to white, corporate funders was an intentional part of the design, one approved by those same white corporate funders. The final report was written as a multi-vocal, multicultural presentation that presented different experiences with and perceptions of the program rather than reaching singular conclusions. The medium of the report carried the message that multiple voices needed to be heard and valued as a manifestation of diversity (Stockdill, et al., 1992). The findings were used for both formative and summative purposes, but the parents and many of the staff were most interested in using the evaluation processes to make themselves heard by those in power. *Being heard was an end in itself quite separate from use of findings.*

Wadsworth (1995, p. 9) has reported that evaluation processes can facilitate interactions between service-providers and service-users in a way that leads to "connectedness" and "dialogue across difference." Each learns to see the service through the others' eyes. In the process, what began as opposing groups with opposing truths is transformed into "an affinity-based community of inquiry" with shared truths.

Certain kinds of evaluation processes, especially participatory and empowering ways of conducting evaluations (Dugan, 1995; Fetterman, 1994, 1995; Fetterman, et al., 1995; King, 1995; Wadsworth, 1995) are also aimed, in some cases, at increasing the effectiveness of the program through the evaluation process rather than just the findings. In such cases, the evaluation process becomes part of the program intervention. The traditional separation between

evaluation and programming falls away and the evaluation is conducted in such a way as to support and reinforce the program intervention. This kind of evaluation design has the purpose of directly enhancing the program's impact by the process of engagement rather than the use of findings. I would submit that such a conceptualization deserves classification beyond the formative-summative distinction. For examples, see Patton (1997, 1988).

Outside the Formative-Summative Boundaries

In summary, I've identified three arenas of evaluation practice in which the formative-summative dichotomy is, for me, inadequate: knowledge-generating evaluations aimed at conceptual rather than instrumental use; developmental evaluation; and use of evaluation processes to support interventions or empower participants. Scriven will argue that none of these are purely evaluative activities. That's precisely the point. He has a tightly and narrowly defined view of evaluation. I have a broad and expanding view. To highlight those differences in definition and perception, let me react to what Scriven considers some typical "fallacies" with regard to evaluation practice since my own practice exemplifies many of these "fallacies."

Logical Fallacies or Divergent Views?

Scriven's 1991 article is organized around ten fallacies that he finds common in evaluation. He uses the term fallacy "in the sense of 'seductive error'" (Scriven, p. 20). The connotation, however, is that those who disagree have lapsed into fundamental errors in thinking. I found this assertion distressing in its intolerance for divergent views and practices. I take issue with Scriven because my experiences and values are different, not, I believe, because my reasoning is fallacious. This point, it seems to me, is critical, because it addresses the diversity of the field of evaluation. These are matters about which evaluators can disagree and the field benefits from discussion of such differences of perspective. Let me illustrate.

Scriven argues that a fundamental and inescapable duty of an evaluator is to render judgment about merit or worth. He dismisses as fallacious the perspective that the use of evaluation is increased if the evaluator plays a facilitating role in helping decision makers and information users arrive at their own judgments. Judge for yourself his tone:

Even amongst professional program evaluators, it is still common to try to avoid adopting any actual evaluation stance although they still call what they do evaluation. This approach is referred to here as "pseudoevaluative investigation," and it results in a description masquerading as evaluation. It is sometimes rationalized by appeal to the following claim:

The Fifth Fallacy (F5) is the belief that the professional evaluator's duty is to give clients the facts and let them assemble (interpret) these according to their own values or to give them the subevaluations and let them put these together.

The first part of this fallacy creates the curious picture of the professional evaluator doing everything except what is normally called evaluating something. In reality, the situation is even worse.... Thus, balking at the last step—the overall evaluation—is like deciding you want to be a virgin after the orgy but before the Day of Judgment. The F5 adherent is nearly always guilty of inconsistency as well as misleading advertising (Scriven, 1991, p. 31).

Scriven rejects and ridicules as "pseudoevaluative" what has become a major form of professional practice, what I have called utilization-focus evaluation (Patton, 1986) and others

call participatory research (e.g., Cousins & Earl, 1995), or collaborative evaluation (e.g., Cousins, et al., 1995). In utilization-focused, participatory, and collaborative approaches, the evaluator, *for the purpose of increasing use*, facilitates the judgments of those who bear responsibility for making decisions. That is, a utilization-focused evaluator facilitates judgment and decision-making by *clearly identified, primary intended users* rather than acting as a distant, independent judge. Since no evaluation can be value-free, utilization-focused evaluation answers the question of whose values will frame the evaluation by working with clearly identified, primary intended users who have responsibility to apply evaluation findings and implement recommendations. In essence, I argue, evaluation use is too important to be left to evaluators.

While concern about utility drives a utilization-focused evaluation, the evaluator must also attend to the evaluation's accuracy, feasibility and propriety (Joint Committee on Standards, 1994). Moreover, as a professional, the evaluator has a responsibility to act in accordance with the profession's adopted principles of conducting systematic, data-based inquiries; performing competently; ensuring the honesty and integrity of the entire evaluation process; respecting the people involved in and affected by the evaluation; and being sensitive to the diversity of interests and values that may be related to the general and public welfare (AEA Principles: Shadish, et al., 1995).

A psychology of use undergirds and informs utilization-focused evaluation. In essence, our research and experience indicate that intended users are more likely to actually use evaluations if they understand and feel ownership of the evaluation process and findings; they are more likely to understand and feel ownership if they've been actively involved; and by actively involving primary intended users, the evaluator is training users in use, preparing the groundwork for use, and reinforcing the intended utility of the evaluation every step along the way.

I understand that there are important times when evaluators are asked by their clients to render independent, summative judgment about the merit or worth of a program. When that is the shared understanding, the evaluator has an absolute obligation to provide such judgment. In other cases, the evaluator and intended users agree that the evaluator will provide information to support decision-making and facilitate judgments by those who carry that responsibility. Naturally, I take umbrage at labeling this approach "pseudoevaluative."

This is no small matter because Scriven goes on to condemn such practices as "the abrogation of the professional responsibility of the evaluator...." (Scriven 1991, p. 32). Thus, he has stated unambiguously that evaluators, like myself, who facilitate the judgments of others in lieu of asserting their own judgments as primary fail to meet the standards of the profession as he sees them. I find such a view sadly restrictive. If widely held, it would be an obstacle to the flowering of a rich and diverse field of practice, which is what professional evaluation is and ought to be. Nor do I believe that the standards of evaluation (Joint Committee), or the principles of evaluation (American Evaluation Association), support the charge of abrogating professional responsibility.

"One of the Most Dangerous ... Mistakes"

Scriven formulates a distinction between "*analytic* evaluation versus *fragmentary* evaluation" and argues that for both formative and summative evaluation, it is ultimately the evaluator's responsibility to judge the entire program. He uses the term "fragmentary evaluation" to refer to the practice of looking at some part of a program and, for example, focusing on

ways of improving that part without evaluating the entire program. He finds such a practice fallacious:

The Sixth Fallacy (F6) maintains that in *formative evaluation*, it is only necessary to point out various respects in which improvement is needed. It is not necessary to draw any overall conclusion.

This is often advanced as *tenable* even if F5 cannot be maintained for summative evaluation. Does it really do any harm? The answer is that it is one of the most dangerous of our agenda of mistakes“ (Scriven, 1991, p. 36).

Later Scriven continues: “Of course, the ‘friendly formative’ image is facilitated by a piecemeal approach; there are no threatening overall conclusions, no deep diagnoses. It’s just that improvement is poorly served, and that’s what formative evaluation is supposed to be all about” (Scriven, 1991, p. 37).

In essence, Scriven elevates one kind of evaluation into the *only* valuable evaluation product—what he calls complete, global and analytical evaluation that puts together the various parts of the program in a way that allows the evaluator to render an overall, comprehensive judgment of merit or worth. He labels other evaluation approaches “fragmentary,” “incomplete,” and “piecemeal”—hardly neutral, descriptive terms. I prefer the term “focused” to describe evaluations in which my clients seek my assistance to improve a particular component of a program. I do such work often. It is valuable and useful work. It is cost effective. Such focus comes from negotiating with primary intended users and clients about how to make best use of limited and scarce evaluation resources. In many cases, the overall effectiveness of the entire program is not at issue. No summative decision is at issue. In a developmental context, program decision makers often need information about a particular component of a program and ways of improving it, and there is nothing inappropriate about evaluators examining that particular component as long as the limitations are understood and accepted by intended users.

Psychological, Social, and Political Considerations

In a major section of his article, Scriven addresses psychological, social and political considerations. Much of his discussion concerns the resistance of evaluation clients to negative findings and the difficulty evaluators have—psychologically—providing negative feedback. He discusses at length the responsibility of evaluators to be hard-nosed and uncompromising in reporting negative results. He asserts that “the main reason that evaluators avoid negative conclusions is that they haven’t the courage for it ...” (Scriven, 1991, p. 42). He pokes fun at suggestions that there might be a “kinder, gentler approach” to evaluation (p. 39).

By way of offering a contrasting perspective, let me share my own experience. First, what takes courage varies in time and place. An internal government evaluator recently told me that what takes courage these days is to conclude that a program is effective and deserving of support and expansion. Trashing programs, on the other hand, is quite welcome from policy makers eager to make headlines through budget cuts.

Second, I work with clients who are hungry for quality information to improve programs. They are people of great competence and integrity who are able to use and balance both positive and negative information to make informed decisions. I take it as part of my responsibility to work with them in ways that they can hear the results, both positive and negative, and use them for intended purposes. I don’t find them resistant. I find them eager to get

quality information that they can use to develop the programs to which they have dedicated their energies. I operate in a way that I hope is constructive and respectful. I try to render judgments, when we have negotiated my taking that role, in ways that can be heard, and I work with intended users to facilitate their arriving at their own conclusions. They are often harsher on themselves than I would be.

It is not so much that it takes courage to provide negative results as it does skill. Nor does it take some special attributes of users for them to be able to receive negative feedback as much as it takes skilled facilitation in recognizing the value of such feedback for long-run effectiveness. I have followed in the tracks of, and cleaned up the messes left by, evaluators who took pride in their courageous, hard-hitting negative feedback. They patted themselves on the back for their virtues and went away complaining about program resistance and hostility. I watched them in action. They were arrogant, insensitive and utterly unskilled in facilitating feedback as a learning experience. I don't aspire to practice such an approach or train students in so practicing.

Ad Hominem Arguments

I confess that Scriven's tone and choice of words made me feel defensive. Though my work was not referenced directly anywhere in the article, I felt attacked because I engage in the practices Scriven disdains. I want to admit to and explore those feelings because, I believe, they may help illustrate my point about varying styles of giving evaluative feedback. Consider the following incident as data. During the panel session at the 1995 International Evaluation Conference that led to this article, I labeled the hard-hitting, let-the-chips-fall-where-they-may, feelings-be-damned style of judgment favored by Scriven as "macho." In his response he took umbrage at being labeled a "male macho pig." He added "male" and "pig." I said "macho." He heard "pig." He made it clear that he rejected that label.

The panel exchange became, from my perspective, ad hominem. I inadvertently made an insensitive reference to a personal tragedy Scriven had recently experienced. I embarrassed myself terribly. I later apologized and he graciously dismissed the matter as insignificant. But it was significant to me because it was part of a reaction against feeling personally attacked. I felt demeaned and insulted by having my approach called cowardly and labeled "pseudoevaluative." I was hurt by the accusation of abrogating my professional responsibility. I was angry at having my thinking labeled fallacious. The article's language *felt* personal and deeply judgmental, even though none of it was directed at me personally. I had trouble separating substantive arguments from the article's condemnatory tone. I think I would have reacted quite differently to a mutually respectful style of collegial discourse, the kind I have always experienced from Michael Scriven personally.

As in any form of feedback, written or verbal, it's hard to hear the substance when the tone is highly judgmental and demeaning. This is true in interactions between parents and children (in either direction), between lovers and spouses (not to imply that these are necessarily different categories), between colleagues, and, most decidedly, between evaluators and intended users. It's true for me, and I consider myself quite thick-skinned. But Scriven's judgments got to me. And I didn't like them. The tone of his arguments affected my openness to his conclusions and my ability to appreciate his point of view. This despite the fact that I admire and respect him greatly.

It's from these kinds of experiences that I have developed a preference for kinder, gentler feedback. Not out of cowardice, but out of a commitment to effectiveness and utility. It strikes

me as terribly important to provide affirmation for those evaluators, whether male or of another persuasion, who are working at overcoming socialized tendencies toward a macho style of evaluative feedback. Being kinder and gentler in an effort to be heard may engender school yard teasing about sissified behavior. (Scriven's accusation of cowardice sent me regressing back to a third grade fight I was ridiculed for walking away from.) But, in a world of diversity, sensitivity and respect are not only brave, they're more effective.

A Menu of Evaluation Purposes

Scriven has done the field of evaluation enormous service by introducing a number of conceptual distinctions. His commitment to the field and dedication to assuring that evaluation practice is of the highest quality helps us all aspire to standards of excellence of which we can be jointly proud. I want to recognize Scriven's enormous contributions, including the important distinction between formative and summative evaluation, even as I reject that distinction as sufficiently all-inclusive to cover the entire, diverse and still-developing world of evaluation.

The field of professional evaluation practice has become too rich and varied to be pigeon-holed into a single dichotomy. I find that responding usefully to a great diversity of situations and client needs requires a large menu of options. Stephen Brookfield's findings on what it takes to become a reflective practitioner seem relevant here. He found that learning to be reflective typically involves a "loss of innocence" that comes from moving away from "dualistic certainty to dialectical and multiplistic modes of reasoning" (Brookfield, 1994, p. 203). In terms of systems theory, it means moving from closed to open systems. The original formative-summative formulation thirty years ago was helpful in evaluation's infancy, but as a maturing field of professional practice, it's time to deal with the complexities of open systems, multiple purposes, and diverse values. Judgment-oriented, comprehensive, and summative evaluation is not the apex or Holy Grail, it is not all of evaluation, and not all formative evaluations are getting ready for summative evaluation.

Evaluation serves other purposes including, but not limited to, the following:

- generating general knowledge about and principles of program effectiveness;
- developing programs and organizations;
- focusing management efforts;
- creating learning organizations;
- empowering participants;
- directly supporting and enhancing program interventions (by fully integrating evaluation into the intervention); and
- stimulating critical reflection on the path to more enlightened practice.

The world of evaluation is vast and rich, and is becoming more vast and richer each year. The evaluation universe, like the physical one, is still expanding.

REFERENCES

- Alkin, M. C., Dailak, R., & White, P. (1979). *Using evaluations: Does evaluation make a difference?* Beverly Hills, CA: Sage.

- Aubel, J. (1993). *Participatory program evaluation: A manual for involving stakeholders in the evaluation process*. Dakar, Senegal: Catholic Relief Services under a U.S. AID grant.
- Barley, Z. A., & Jenness, M. (1993). Cluster evaluation: A method to strengthen evaluation in smaller programs with similar purposes. *Evaluation Practice*, 14(2), 141–147.
- Brookfield, S. (1994). Tales from the dark side: A phenomenography of adult critical reflection. *International Journal of Lifelong Education*, 13(3), 203–216.
- Campbell, J. L. (1994). Issues of cluster evaluation use. Paper presented at the 1994 Annual Meeting of the American Evaluation Association, Boston, MA.
- Chen, H. (Ed.) (1989). Special issue: The theory-driven perspective. *Evaluation and Program Planning*, 12, 4.
- Chen, H. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Chen, H., & Rossi, P. (1987). The theory-driven approach to validity. *Evaluation and Program Planning*, 10, 95–103.
- Council on Foundations. (1993). *Evaluation for foundations: Concepts, cases, guidelines and resources*. San Francisco, CA: Jossey-Bass.
- Cousins, J. B., Donohue, J. J., & Bloom, G. (1995). *Collaborative evaluation: Survey of practice in North America*. Unpublished paper presented at the International Evaluation Conference, Vancouver, Canada. Monograph inquiries: bcousins@educ-1.edu.uottawa.ca.
- Cousins, J. B., & Earl, L. (Eds.) (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London: Falmer Press.
- Dugan, M. (1995). Participatory and empowerment evaluation: Lessons learned in training and technical assistance. In D. M. Fetterman, A. J. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Newbury Park, CA: Sage.
- Fetterman, D. M. (1994). Empowerment evaluation. American Evaluation Association Presidential Address. *Evaluation Practice*, 15(1), 1–15.
- Fetterman, D. M. (1995). Steps of empowerment evaluation: From California to Cape Town. *Evaluation and Program Planning*, 17(3), 305–313.
- Fetterman, D. M., Kaftarian, A. J., & Wandersman, A. (Eds.) (1995). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Newbury Park, CA: Sage.
- General Accounting Office. (1992). *Adolescent drug use prevention: Common features of promising community programs*. GAO/PEMD-92-2. Washington, DC: U.S. GAO.
- Guttman, D. & Susman, M. B. (Eds.) (1995). Exemplary social intervention programs for members and their families. Special issue of *Marriage and Family Review*, 21, 1–2.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards*. Thousand Oaks, CA: Sage.
- Kellogg Foundation, n.d. (circa 1995). W. K. Kellogg Foundation cluster evaluation model of evolving practices. Battle Creek, MI: Kellogg Foundation.
- King, J. A. (1995). Involving practitioners in evaluation studies: How viable is collaborative evaluation in schools. In J. B. Cousins, & L. Earl (Eds.) *Participatory evaluation in education: Studies in evaluation use and organizational learning*, pp. 86–102. London: Falmer Press.
- Leviton, L. A., & Hughes, E. F. X. (1981). Research on utilization of evaluations: A review and synthesis. *Evaluation Review*, 5(4), 525–548.
- Mayer, S. E. (1993). Common barriers to effectiveness in the independent sector. In *A Vision of Evaluation*, pp. 7–11. Washington, D.C.: Independent Sector.
- Mayer, S. E. (1994). *Building community capacity: The potential of community foundations*. Minneapolis, MN: Rainbow Research, Inc.
- Patton, M. Q. (1986). *Utilization-focused evaluation* (2nd ed.). Newbury Park, CA: Sage.
- Patton, M. Q. (1988). Integrating evaluation into a program for increased utility and cost-effectiveness. In J. A. McLaughlin, L. J. Weber, R. Covert, R. B. Ingle (Eds.) *Evaluation Utilization*, New Directions in Program Evaluation Series, No. 39, 85–95. Jossey-Bass, San Francisco.
- Patton, M. Q. (1994). Developmental evaluation. *Evaluation Practice*, 15(3), 311–320.

- Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). Newbury Park, CA: Sage.
- Patton, M. Q., Bringewatt, M., Campbell, J., Dewar, T., & Mueller, M. (1993). *The McKnight Foundation Aid to Families in Poverty Initiative: A synthesis of themes, patterns and lessons learned*. Minneapolis, MN: The McKnight Foundation.
- Patton, M. Q., Grimes, P. S., Guthrie, K. M., Brennan, N. J., French, B. D., & Blyth, D. A. (1977). In search of impact: An analysis of the utilization of federal health evaluation research. In C. Weiss (Ed.) *Using social research in public policy making*, pp. 141–164. Lexington, MA: D.C. Heath.
- Rossi, P. H., & Freeman, H. E. (1985). *Evaluation: A systematic approach*. Beverly Hills, CA: Sage.
- Sanders, J. R. (1994). Methodological issues in cluster evaluation. Paper presented at the Annual Meeting of the American Evaluation Association, Boston, MA.
- Schorr, L. (1988). *Within our reach: Breaking the cycle of disadvantage*. New York: Doubleday.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven, M. (Eds.) *Perspectives of curriculum evaluation*, pp. 39–83. AERA Monograph Series on Curriculum Evaluation, 1. Chicago: Rand McNally.
- Scriven, M. (1991). Beyond formative and summative evaluation. In M. W. McLaughlin & D. C. Phillips (Eds.) *Evaluation and Education: At Quarter Century*, pp. 18–64. Chicago, IL: The University of Chicago Press.
- Scriven, M. (1994). The final synthesis. *Evaluation Practice*, 15(3), 367–382.
- Scheirer, M. A. (1987). Program theory and implementation theory: Implications for evaluators. In L. Bickman (Ed.) *Using Program Theory in Evaluation*, New Directions for Program Evaluation Series, No. 33, 59–76. Jossey-Bass, San Francisco.
- Shadish, W. R., Jr. (1987). Program micro- and macrotheories: A guide for social change. In L. Bickman (Ed.), *Using Program Theory in Evaluation*, New Directions for Program Evaluation Series, No. 33, 93–110. Jossey-Bass, San Francisco.
- Shadish, W. R., Jr., Newman, D. L., Scheirer, M. A., & Wye, C. (1995). *Guiding principles for evaluators*. New Directions for Program Evaluation Series, No. 66. San Francisco: Jossey-Bass.
- Stockdill, S. H., Duhon-Sells, R. M., Olson, R. A., & Patton, M. Q. (1992). Voices in the design and evaluation of a multicultural education program: A developmental approach. In A. M. Madison (Ed.) *Minority issues in program evaluation*. New Directions in Program Evaluation Series, No. 53, 17–34. San Francisco, CA: Jossey-Bass.
- Wadsworth, Y. (1995). "Building In" research and evaluation to human services. Unpublished report to the Winston Churchill Memorial Trust of Australia, Melbourne.
- Wargo, M. J. (1989). Characteristics of successful program evaluations. In J. S. Wholey & K. E. Newcomer (Eds.) *Improving government performance: Evaluation strategies for strengthening public agencies and programs*, pp. 71–82. San Francisco, CA: Jossey-Bass.
- Weiss, C. H. (1990). Evaluation for decisions: Is anybody there? Does anybody care? In M. Alkin (Ed.) *Debates on evaluation*, pp. 171–184. Newbury Park, CA: Sage.
- Weiss, H. B. Greene, J. C. (1992). An empowerment partnership for family support and education programs and evaluations. *Family Science Review*, 5(1,2), 145–163.
- Worthen, B. R. (1994). Conceptual challenges confronting cluster evaluation. Paper presented at the Annual Meeting of the American Evaluation Association, Boston, MA.