*Articles should deal with topics applicable to the broad field of program evaluation. Articles may deal with evaluation practice or theory, but if the latter, implications for practicing evaluators should be clearly identified. Examples of contributions include, but are not limited to, reviews of new developments in evaluation and descriptions of a current evaluation effort, problem, or technique. Results of empirical evaluation studies will be published only if the methods and/or findings have clear utility to other evaluation practitioners. Manuscripts should include appropriate references and normally should not exceed 10 double-spaced typewritten pages in length; longer articles will occasionally be published, but only where their importance to AJE readers is judged to be quite high.*

Have We Learned Anything New About the Use of Evaluation?

CAROL H. WEISS

ABSTRACT

"Use of evaluation" used to mean the use of results for making program decisions. Now we are aware of its larger dimensions. Many aspects of evaluation can be used—ideas and insights, signals that are sent by the very fact of evaluating, the processes of learning through cooperation with evaluation, the choice of outcome indicators that herald success, even the study design. "Use" encompasses a broad array of effects by multiple classes of users. The paper highlights the importance of considering further potential collective users—the program organization, client groups, and civil society.
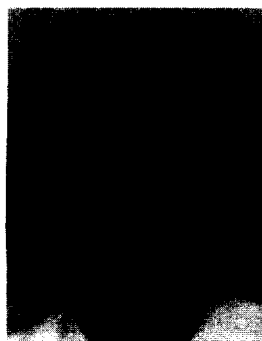
Carol H. Weiss

Carol H. Weiss • Harvard Graduate School of Education, Harvard University, Cambridge, MA 02138.

INTRODUCTION

Ten years ago, in the plenary address at the American Evaluation Association, I discussed the uses of evaluation (Weiss, 1988a). The speech, which I thought was a fair summary of the state of the field, caused a little ruckus, and it took AEA ten years to invite me back again. This time I am asked to talk about theories of use, which sounds less incendiary than experience with use, and I have been urged by AEA's President, Will Shadish, not to make this a talk of advocacy. I intend to follow both injunctions, but perhaps I will still cause a bit of a stir.

My earlier address maintained that evaluators should do top-quality evaluations that are responsive to the concerns of those whose programs they evaluate, and they should do their best to communicate results in multiple ways to ensure that program people know the results and see their relevance for program action. I stressed working with program people at the start to understand their concerns and again at the end to help interpret the implications of the results.

But I also held that evaluators should not be held accountable for failures to use their results. Even when program staff know about the findings, understand them, believe them, and see their implications for improving the program, many factors can interfere with their using results for program improvement. Among the possible obstacles are conflicting beliefs within the program organization, with staff unable to agree on what the important issues are; conflicting interests between program units, so that there is a squabble over resources or clout; changes in staff, so that new people coming in have different priorities from those in force when the evaluation was begun; rigidity of organizational rules and standard operating procedures that prevent adoption of improved strategies suggested by the evaluation; shifts in external conditions, such as budget cuts or changes in the political climate that make the organization unable to respond to the need for change revealed by evaluation, and so on. I have experienced a number of these kinds of obstacles, and I did not feel responsible when findings were not used. So while I thought it was reasonable to ask evaluators to work hard at communicating their findings, I also counseled that they should not hold out unrealistic expectations for use.

Michael Patton was upset with my talk. He thought it was a defeatist message. He leaped up on the platform after I finished and said he disagreed with me and invited the audience to his session the following day. He believed and still believes that evaluators who work collaboratively with program stakeholders can get their results used. The new edition of Patton's (1997) book, *Utilization-Focused Evaluation*, gives directions for how such participatory strategies can yield use—which he defines as intended use by intended users. Working with program staff enlists their attention, knowledge, and commitment to using the results.

I wrote a response, agreeing with many of Patton's points at the program level but claiming a broader stage for evaluation. All three papers were published in this journal (Weiss, 1988a, 1988b, Patton 1988). They were reprinted in Alkin's (1990) book, *Debates on Evaluation*, along with comments by him and others about the reasons for the disagreement.

I am not going to re-hash the old arguments here. (Well, only a little.) As much fun as it might be for the speakers to continue expatiating their ideas and scoring points, it is apt to be deadly dull for the audience. I want to go on to talk about some of the new ideas that have entered the field of knowledge use in the last decade. For example, a good review of the utilization literature now exists. Shulha and Cousins (1996) gave a talk on this topic at these meetings last year, and an abbreviated version of their paper appears in the most recent issue of this

journal (Shulha & Cousins, 1997). I recommend it to you, so that I do not have to go back and review what a lot of you have been reading or, in some cases, writing. Rather, I will introduce a few ideas from the recent literature that have captured my fancy.

But first let me answer the question that the title raises: Have we learned anything new about the use of evaluation? I probably should not give away the denouement so early, but this is not a mystery story, so I will. My answer is: yes we have learned some things, but the learnings have come more from applying new constructs and perspectives than from research on evaluation use.

We used to do empirical studies to identify the correlates of use; we studied characteristics of studies, characteristics of potential users, and communication strategies that were associated with greater use of results. But we have come to a growing realization of how complicated the phenomenon of use is and how different situations and agencies can be from one another. We are also aware of the higher-order interactions among characteristics of the study, the evaluator, the setting, and the communication methods used. Because of these complexities, it is conceptually and theoretically difficult to reduce the elements to a set of quantitative variables. The old kinds of studies that tried to isolate the few keys to utilization have largely gone out of style. What we have learned in the past decade has come about largely through experience. Fortunately, a number of hardy souls have written about their experience, so we can all profit from it. Collaboration between evaluators and program staff all through the evaluation process tends to increase the local use of evaluation findings. Not always; there are still many potholes along the path. But participatory methods do seem to help. They have their drawbacks as well. I will come back to this.

Probably more important than empirical research for addressing the question of use in the last ten years has been the inheritance of new perspectives from other fields. If our understanding of the use of evaluations has advanced, it is partly because we have new ways of thinking about it. As Rivlin (1971) said about the use of policy analysis, it may not have solved any problems, but it has led to a better class of problems. So it is with our thinking about evaluation use. We have not solved the problem but we are thinking about it in more interesting ways. I am going to talk about how some new perspectives have influenced our understanding of use, or in Shadish's (1997) term, what their implications are for a theory of use.

This discussion comes in three parts. The first is a brief discussion about what we mean by use. The second is a discussion of what it is from evaluation that is used. In the third, I talk about who the users are. Finally, I will end with a few suggestions about how we can improve the use of evaluation.

What Do We Mean By Use?

When we first began to think about evaluation use, we meant use for decision making. We expected evaluation to produce findings that would influence what program and policy people decided to do next. They might end a program, extend it, modify its activities, or change the training of staff; they were expected to use what evaluators had found in order to make wiser decisions. This kind of use has come to be known as instrumental use.

Instrumental use for decision making is fairly common when the evaluator does a good job of understanding the program and its issues, conducting the study, and communicating results. At least instrumental use is common under three conditions: (1) if the implications of the findings are relatively non-controversial, neither provoking rifts in the organization nor

running into conflicting interests, (2) if the changes that are implied are within the program's existing repertoire and are relatively small-scale, and (3) if the environment of the program is relatively stable, without big changes in leadership, budget, types of clients served, or public support. There is a fourth situation that encourages use, too—when the program is in a crisis or paralysis, and nobody knows what to do. *In extremis*, they may turn to evaluation. So those are examples of the first kind of use—instrumental use for making decisions.

A second kind of use is conceptual use by the local program people. Even if they are blocked from applying the findings to decisions at the time the study is reported, the findings can change their understanding of what the program is and does. They gain new ideas and insights. If they have been engaged in the process of evaluation, they learn even more about strengths and weaknesses and possible directions for action. When organizational conditions become favorable, they can then use their new conceptual understandings in instrumental ways.

A third kind of use is to mobilize support for a position that people already hold about the changes needed in the program. Often program managers and operators know what is wrong and what they need to do to fix the problem. They use evaluation to legitimate their position and gain adherents. Evaluation becomes an instrument of persuasion.

The fourth and final kind of use is influence on *other* institutions and events—beyond the program being studied. As hundreds and thousands of evaluations are done on similar kinds of programs, evidence mounts up about Head Start or pregnancy prevention or rehabilitation of prison inmates. Sometimes the evidence is synthesized through meta-analysis, which gives an overall estimate of the outcomes of programs of similar kind; sometimes the synthesis comes through qualitative reviews. Often the aggregation of conclusions is more haphazard. But evaluation evidence often comes into currency through professionals and academics and evaluators, and it influences networks of practicing professionals and policy wonks (Heclo, 1978), infiltrates advocacy coalitions (Sabatier, 1993), alters policy paradigms (Weiss, 1980), changes the policy agenda (Kingdon, 1995), and affects belief communities within institutions (Radaelli, 1995). Such influences are by no means commonplace or easy, but knowledge is one of the important forces in policy and program formulation. When evaluation adds to the accumulation of knowledge, it can contribute to large-scale shifts in thinking—and sometimes, ultimately, to shifts in action.

Not all evaluators aspire to such kinds of influence. Nevertheless, enough cases have been studied to show that the "enlightenment" kind of use is not a negligible category.

What Is It That Is Used?

First, of course, is the evaluation findings. Originally, all we thought about was findings. I wrote the first paper ever published on evaluation utilization (Weiss, 1967), and that is what I talked about. Evaluation produced findings about program processes and program outcomes, and evaluators expected users to apply the findings to their programs. If the evaluator developed recommendations based on the findings, then the recommendations were supposed to be used, too. In fact, it was easier to study the use of recommendations than anything else, so studies of evaluation often counted up the number of recommendations that had been implemented (Leviton & Boruch, 1983).[1]

Two, we realized that policy and program people can use the ideas and generalizations from evaluation, even if they do not apply specific findings. So, for example, the generalization that a neighborhood agency does not sustain changes in its patterns of service if it is the

only agency making the change has led to plans for coordination across agencies. I have written extensively about the use of ideas and insights—the enlightenment that comes from evaluation.

A third aspect of evaluation that can be used is the sheer fact that an evaluation is being done. On the bad side, the fact of evaluating can lead a program director—like one man I worked with—to fend off demands for needed change. He used the excuse that an evaluation was in progress and he had to wait until the evaluation was finished before he could do anything at all. Evaluation became an excuse for inaction. Less obstructive but not much more constructive, the sheer fact of evaluation can be used to demonstrate the program manager's rationality and good management style (Feldman & March, 1981). Here evaluation becomes a surrogate for good management. Perhaps the manager will go on and attend to the findings and improve the direction of the program. But sometimes he is content to point to the very fact of evaluating as both signal and symbol of his progressive mode of work.

In similar vein, some program managers use the existence of evaluation as a device of accountability. That is, they show that they are being accountable by pointing to the evaluation in progress. Eventually, this may become a responsible position by informing higher levels of the agency and the public about what is going on. But when the sheer existence of evaluation is used as a display of accountability, it can substitute illusion for reality.

Perhaps more important, the fact that an evaluator agrees to undertake the study helps to legitimate the program. The evaluator is essentially sending a message that the program is a serious effort, worth the time and resources that it takes to evaluate. Something can be learned from the program, and evaluation professionals are willing to engage themselves in its study. The evaluation provides an aura of substance and legitimacy.

At the same time, the willingness of evaluators to subject the program to study also suggests that there is something problematic about it. It may not be obtaining the desired results. It needs to be subjected to scrutiny. The sheer fact of evaluating sends a message that the program still has to win its spurs. It may need to be modified.

Outside audiences hear that double-barreled message. It influences their image of the program, even before any data or findings become available. Depending on their orientation, they pay more attention to the legitimation of the program or to the uncertainty of its worth.

On the positive side, the process of evaluating can have happy effects on future program events. As Greene (1988) and others have written, program staff who participate in defining and framing the evaluation begin to think more deeply about what they are trying to accomplish. They talk to their colleagues and try to reach consensus. Such discourse can have the effect of clarifying the program's goals, and once agreement is reached on these matters, it can increase the cohesiveness of managers and staff. People are now working on the same things and trying to attain the same ends. Participating in the evaluation helps staff to remember why they are doing this job, and it can reinvigorate their practice.

Collaborative evaluation has the side benefit of helping program people reflect on their practice, think critically, and ask questions about why the program operates as it does. They learn something of the evaluative cast of mind—the skeptical questioning point of view, the perspective of the reflective practitioner. Such a perspective can stand them in good stead in their practice. It might even encourage them to set up information systems that will outlive the study in progress and allow them to review their progress on a continuous basis.

When the evaluation is a theory-based evaluation, staff have to examine their assumptions about how the program works. Thinking though their expectations can show them which expectations are wild leaps of faith and need to be re-thought and which need to be supported

by additional activities. For example, if a school program expects that getting parents to talk to teachers will lead to improvements in students' school work, they come to realize that a parent-teacher conference is unlikely to achieve that goal unless auxiliary steps are taken. When participatory evaluation yields consensus about assumptions and theories, the payoff to the program is likely to be high—staff pulling together to achieve the same steps en route to long-term objectives.

Another way that the procedures of evaluation can influence action is through the choice of what to study. When the staff knows that the focus of evaluation is high school graduation rates, there is a temptation to push that indicator higher, even if teachers have to graduate students who do not deserve to graduate. Whatever is studied or measured tends to get priority attention, at least if upper echelons are likely to pay attention and dispense punishments and rewards. The phenomenon is common enough to have a well-known label: teaching to the test. Teaching to the test can be an unfortunate consequence of evaluation if teachers stop teaching the other important things in the curriculum and devote all their energies to teaching the skills that will be tested. More generally, the danger is that other aspects of the program will be ignored in order to concentrate on the two or three items that show up as indicators of program success—arrest rates, job placement rates, and so on.

But teaching to the test can also have good consequences. In fact, some education reformers are actually using testing as a means to bring about change in classrooms. They are highlighting things that they want to matter. For example, they put into tests questions that call for critical thinking in order to sway teachers' classroom behavior. They expect that teachers will teach critical thinking skills in their classes because they want their students to do well on the test.

The choice of measures (or in qualitative evaluations, the focus of study) can influence program operations. That is a way that evaluation is used. And not only the measures but also the design of evaluation itself can be "used." It may sound far-fetched that the very methods of study can have influence on subsequent program and policy events, but Breslau (in press) has written a fascinating account of how this has happened in the federal employment and training field. He shows how the use of evaluation was based not only on the transmission of results but on the evaluators' categories of data, design, and analysis.

The manpower development and training act of 1962 created a federal office to provide funds to local training programs to assist chronically unemployed workers. The local training providers were largely the pre-existing state employment services and vocational educators. The federal staff wanted to use the program to direct attention to serving the underprivileged, mainly urban minority groups. Local providers, on the other hand, wanted to maintain established relationships with local businesses, labor unions, and local government. Data became an object of contention. Local staffs evaluated their success by the number of placements made and the gains in trainees' earnings. These data showed considerable success. The federal staff insisted on comparisons between the people served and the total population of unemployed, and thus found that local providers were often serving the most skilled and easiest to employ, a practice that came to be called "creaming." In the federal view, such practices were noted as a failure.

Furthermore, design became an issue. Local providers used before-and-after evaluation designs to demonstrate the effectiveness of their training, and the data showed that trainees achieved sizable gains in earnings. The federal staff required net impact evaluation—a study of the difference between what the trainees earned and what they would have earned in the absence of the program. They insisted that evaluations be done by random assignment or by

statistical methods that approximated random assignment. Conversely, to local providers, net impact did not make much sense; it was divorced from any real-live people. What they knew was whether individuals were placed in jobs and what their gain in earnings was, but this was irrelevant to net impact estimates.

For their part, evaluators and economists highly approved the adoption of net impact evaluation as an advance in method. But many did not realize that the choice of method was a struggle between staffs at different levels, who were using the choice of evaluation method to gain control over the goals and operations of the program. Because the office of management and budget (OMB) had the final say over evaluation contracts, the federal staff won out. Federal employment and training programs and policies were directed at serving the hard-to-place and the autonomy and discretion of local program operators was drastically reduced.

Breslau's example shows the way in which the methods of evaluation can be "used." The insistence on net impact, he says, helped to give federal staff the upper hand. One can imagine similar contention over quantitative versus qualitative approaches. So what looks like a purely technical aspect of evaluation can have extensive influence on policy and program. Thus, the following can all be used: (1) findings or recommendations, (2) ideas and generalizations from evaluation, (3) the very fact of evaluating, which can be used for good or ill, (4) the focus of the study, especially the measures used, and (5) the design of the study.

Who Are The Users?

In the early days, evaluators expected that major users would be program sponsors (the people who paid the program's bills), program directors (who administered the program at national or local levels), and perhaps program practitioners (the staff in direct contact with clients). These were the stakeholders, and evaluators took their lead from them and negotiated the terms of the study with them.

Not too long thereafter, evaluators became uncomfortable with working primarily for the powers-that-be and sought to take the needs and wants of program clients into account. Clients were often the poor, the sick, the old, mentally ill, delinquents, substance abusers, and other groups generally denigrated in society and largely powerless in their dealings with society—and with the program. Evaluators often chose to use evaluation to take clients' values and interests into account and address their concerns. Rather than bolstering the status quo, these evaluators hoped to bring about the kinds of change that made programs more responsive to the needs of participants. They wanted, in Saville Kushner's recent words on Evaltalk, to "challenge the privileged discourses of the powerful." Their stance marked an effort to transform evaluation from a mechanism for sustaining and reinforcing current practice through minor adjustments to a means of remedying inequalities and redistributing power.

Evaluators have also served a wide range of users beyond the immediate program. Evaluation can be used by managers of other similar programs to learn how to improve *their* programs, by federal officials and foundation officers to figure out what to fund or how to improve the operation of programs they sponsor, by policy makers in legislatures to adopt amendments to existing policies or new policies, by social scientists to see what new knowledge has accrued and incorporate it into theory—and into textbooks, and by other evaluators to profit from the findings and the methods of study. Of course, not every study will interest the whole array of possible users, or be worth their attention. But evaluators can envision a wide canvass for their efforts. They can construe as potential users of evaluation a broad array of people with a variety of perspectives.

There is another kind of potential user beyond the individuals mentioned so far—the program organization. Many of us are sociologists, but we have tended to talk about evaluation users as individuals. We know that the individuals function within organizations, but often we lose sight of that fact. One concept that might help is that of the learning organization (Senge, 1990, Argyris & Schon, 1996). The notion of the learning organization is relatively recent, although its roots go far back. The concept has important implications for the construal of evaluation use. Basically, the notion is that outside conditions are changing so rapidly that in order to survive, organizations have to keep learning to adapt. Practices and procedures that worked well in the past are bound to become obsolete, and every organization should keep open to new ways of doing what it does—and even to the thought of doing new things.

Members of a true learning organization have to be interested in evaluation. They will probably work on a collaborative team to design a study and implement it. But where evaluation looks at the past, they are more interested in the future. They want ideas for improvement. Evaluation is good at pinpointing successes and shortfalls and identifying where changes must be made, but it is not always the best source of fresh ideas for tomorrow. People have been writing in the evaluation literature about the necessity for evaluators to take on a broader assignment and become consultants and advisors to organizations and programs. But what are our qualifications for assuming such roles? Our strengths are usually our technical skills, our ability to listen to people's concerns and design a study that addresses them, and the ability to communicate. Are we the best source of new ideas? We are smart, and we could read up on the program area, look at evaluations and meta-analyses, and reflect on our own experience. We can go beyond the data from a specific study and extrapolate and generalize. But does that qualify us as experts? We need to examine our credentials for undertaking this kind of mission. If we want to take it on, we have to learn a great deal more about many things than most of us now know.

The concept of a learning organization alerts us to another important point. Programs are mounted and run in organizations. They are almost never the unique province of free-floating staff who are free of constraints. On the contrary, almost all programs function within rigid limits imposed by law, tradition, procedures, regulations, accustomed habits of doing things, and restraints from other organizations in the interorganizational field. Programs operate within systems of funding, personnel recruitment and promotion, staff inservice training, and so on. To think of increasing the use of evaluation without considering the organizational surround is to miss a good part of the story. If results are to be implemented for program improvement, *organizational conditions* may have to be changed—to remove impediments and to supply supportive structures to incorporate and sustain new approaches and activities. Effective use of evaluation often requires institutional changes to undertake and support new activities and to provide incentives and rewards for staff who adopt the new stance. If changes are to be sustained over time, the ways in which institutions function have to be addressed.

I would like to propose another category of evaluation user: the public, or at least that segment of the public that pays attention to social and political issues, the informed public. In fact, a major audience for evaluation might well be "civil society." Civil society is a term that has come into currency recently to characterize that part of the public that is active in communal and associational life. In the past we often called this group "the community," but the word "community" suffers from vagueness on the one hand and multiple meanings on the other. "Civil society" comprises people in their voluntary and civic engagements (Wolfe, 1997, Solo, 1997).

The reason for adding civil society to the user category is that evaluations of any moment involve issues that go beyond the interests of the people involved in the kinds of programs under study. Broader publics need to hear about the successes and shortfalls of program implementation and program outcomes. Active in many local initiatives, members of civil society can use evaluative information in the program activities in which they are engaged as volunteers, board members, and advisors. More than that, they are opinion leaders in their communities. They can use evaluation to illustrate the successes that programs can have and help to counteract the apathy and hostility that many social programs face these days.

Even when evaluation shows inadequate implementation or small outcome effects, the information is important. When informed and active publics know where program shortfalls are and how other programs have overcome them, they may be apt to take a positive stand toward improving the program rather than gutting it. It is an article of faith in democracy that well-informed publics are more likely to make wise decisions than those without data who can be swayed by emotion and demagoguery. It would be interesting to seek empirical confirmation of the idea that civil society informed through the results of evaluation would be more likely to support program improvement.

CONSTRUCTIVISM—A DIVERTISSEMENT

I have come all this way without mentioning the word "constructivism," let alone "postmodernism." Yet these currents of thought pose a serious challenge to the entire concept of evaluation use. In terms of some constructivist positions, all findings are conditional and contingent. They arise in a particular sociocultural environment and have meaning only in the concrete conditions of this particularity. Furthermore, they can be construed differently by each person, and one person's perspective is as valid as any other's.

In this view, conclusions of any investigation are valid only within the immediate context from which they arose. In addition, they are subject to alternative meaning-making by each individual, who processes the information through his own history, psychology, epistemology, and identity.

Such a view can obliterate the whole idea that evaluation findings can be used – at least used in the sense of being transferred from a specific time and place and program population to other times, places, or people. Strains of constructivist and postmodern thought provide powerful support for using evaluation findings in the local site where the study was done, not in the far-flung locations I have been talking about. They provide powerful support, too, for the involvement of local people in the study, because it is through the ongoing process of discussion over time that meanings can be negotiated and common interpretations reached. The constructivist press is toward thinking locally and acting locally.

I have sympathy with some of these positions. They highlight some of our unreflective assumptions and inanities about using evaluation to find "what works" across the board. But I do not believe that we have to throw out the baby with the bathwater. We can not transfer (and use) evaluation findings mechanically from one place to another. But certainly we can gain important information about what happens in the sites studied, and we can use that information as illustration and metaphor of what can happen under similar conditions elsewhere. Schon (1987) has written eloquently about "reflective transfer," i.e. thinking about the conditions under which the original study was done and analogizing the results to other places when conditions are "similar enough."

Constructivist ideas make us think more carefully about using evaluation results to construct generalizations and syntheses. To that extent, they supply a healthy corrective. But I doubt that they vitiate the use of evaluation—or in fact, the reliance on any social science. Conditions and interpretations differ, but there is enough commonality across people, programs, and organizations to make a functioning social world possible. We are not condemned to a life of studying each event *de novo* in its unique setting. Generalization is a reasonable—and necessary—pursuit.

CONCLUSION

I was asked to talk about the *theory* of evaluation use. If I adhere to Will Shadish's commodious definition of "theory," I think I have made a stab at my assignment.

But let me end on a practical note. I am interested in effective utilization of evaluation—not necessarily more utilization but more effective utilization, use for improving daily program practice and also use for making larger changes in policy and programming.

What do we know about making evaluation more helpful to practitioners, program planners, and policymakers? The best way that we know to date of encouraging use of evaluation is through involving potential users in defining the study and helping to interpret results, and through reporting results to them regularly while the study is in progress. As Huberman (1989) has recommended, such strategies may be even more successful if the evaluator maintains contact with users for a long period after the study ends and if the evaluator listens to what program people have to say about the meaning of evaluation findings and learns from them and with them about how to apply the findings in concrete situations. This kind of "sustained interactivity" transforms one-way reporting into mutual learning.

Participatory styles of evaluation have advantages, but we need to be aware of their limitations as well. One is a tendency to involve only program managers and staff in the definition of the evaluation study and its conduct and interpretation.. Despite talk about the inclusion of all stakeholder groups, participation often narrows down to program staff. I have been reading reports of participatory and empowerment evaluations. They involve program managers, sponsors, board members and staff—primarily professionals—but it seems that only a minority of them involve *clients* in the evaluation process. This choice makes sense. Professionals have the most direct opportunity to use results, and they are also likely to feel personally attached to the program and willing to invest the time in its evaluation. But program *clients* also have a big stake in the program. I wish their inclusion were more widespread, especially clients from marginal groups in society. Difficult as it is to engage them in evaluation activities, they are apt to have different interests and concerns from those of staff, and addressing their questions would broaden the scope of the study. It would also bring people with different orientations and perspectives together to hear and come to understand each other's points of view. Including clients in the evaluative process would also help redress the inequalities in access to influence and the imbalance of power that beset marginalized groups in their social interactions with the larger society. In House's term, it would move—even if only a little way—toward greater social justice (House, 1993).

By focusing on issues that staff can do something about directly, participatory evaluations no doubt increase the likelihood that findings will be used, but the choice also means that the study is apt to be conservative in nature. Program professionals tend to direct evaluation to issues over which they have authority. They function within the existing power structure of

the organization. The study will not be looking into distributional questions, questions of fundamental change, alternative activities, new staff roles—nothing that is apt to rock the boat. Studies that emerge from staff collaboration will not challenge the existing division of labor or lines of authority. Staff will generally want to know relatively small things, such as whether more days of staff development lead to greater teacher competence, or whether staff development is better scheduled during school hours or on weekends. They are not likely to ask whether staff development is sufficient to ensure that teachers implement fundamental school reform.

Participatory evaluation, for all its good points, may have the process upside down. We evaluators are trying to enlist program people in *our* work—doing evaluation—while what they want is to enlist us in *their* work—improving the organization and its programs. They want the best evidence and ideas going.

What they may really want is a forum, a place where program managers, planners, and policymakers can interact with evaluators, researchers and academic experts to discuss *their* questions, offer their own experience and learn about the state of knowledge in the field. The forum would be a place to negotiate the meanings of available knowledge for their own particular circumstances. Similar discussions might involve practitioners and clients. In this kind of action-oriented setting, evaluation would be one source of knowledge and insight—but only one. Still, the forum would provide an opportunity for the use of evaluation conclusions, as well as for acknowledgement of the limits of evaluation conclusions, as inputs into an action-oriented, problem-solving, organizationally specific frame. Evaluators would listen as well as speak, learn as well as teach, and pool their expertise with the expertise of other groups in negotiating at least provisionally useful directions for action.

Reimers and McGinn (1997) have written a book titled *Informed Dialogue* that presents case studies of this kind of forum. They worked with education ministers in developing countries and their vignettes demonstrate how they painfully moved from a tell-them-the-findings approach to an ongoing dialogue about contemplated changes that drew on research and evaluation knowledge as appropriate.

Use is about change. Any theory of evaluation use has to be a theory of change. But it is not change any which-way the evaluator wants. It is not supposed to be change based on the evaluator's ideological or political predilections. That is not why people invited us onto the scene. They assume that the direction for change that comes out of the evaluation will be based on what the evaluator finds out. Listen to that statement: what the evaluator finds out. The formulation has an old-fashioned positivist ring.

We know that what the evaluation finds out is influenced by what it looks at. It is heavily dependent on the questions the evaluator chooses to ask and which elements of the program she chooses to study and which she accepts as given. What evaluation finds out depends on what the evaluator focuses on and measures and how she measures them. It is even affected by the study design she uses.

We now realize that these features of evaluation affect not only the conclusions of the study. We see that evaluation questions, measures, research design, and the sheer fact of evaluating, by themselves send messages—and can be used. The messages are picked up by audiences *in* the program, *around* the program, and at some remove.

Evaluators have to be more self-conscious about the implicit messages as well as the explicit messages they send. We should be wary of thinking about the matter of utilization as solely the province of professionals involved in the program under study. Other audiences will

be paying attention and they may attend to parts of the message we were not aware we were sending.

Evaluators have to be much more aware of our responsibility for the choice of programs to evaluate and for the ways in which we evaluate them. Many of these choices are not ours alone. Program stakeholders—sponsors, staff, clients—have a say in the decisions. But we have to understand that these aspects of evaluation can be used, sometimes in ways we have no inkling of until much, much later.

We cannot control all aspects of the evaluation. But we have a responsibility to think clearly about which elements of the evaluation we want to see used, by whom, and for what purposes. Then we have to go about the business of interpreting, communicating, and disseminating evaluation messages forcefully and expertly *over time* so that important learnings get through. We cannot leave the process of evaluation utilization to chance or regard it solely as an in-house conversation among colleagues. The evaluator has to seek many routes to communication—through conferences, workshops, professional media, mass media, think tanks, clearinghouses, interest groups, policy networks,—whatever it takes to get important findings into circulation. And then we have to keep our fingers crossed that audiences pay attention.

NOTE

1. Of course, not all recommendations are created equal. Some are basically trivial (such as improve program record-keeping), whereas others call for fundamental changes in the way a program is run. To count up the number of recommendations that a program follows can result in pretty meaningless numbers.

REFERENCES

Alkin, M. C. (1990). *Debates on evaluation*. Newbury Park, CA: Sage.

Argyris, C. & Schon, D. A. (1996). *Organizational learning II: Theory, method and practice*. Reading, MA: Addison-Wesley.

Breslau, D. The political power of research methods: knowledge regimes in US labor market policy. *Theory and Society* (In press).

Feldman, M. & March, J. G. (1981). Information in organizations as signal and symbol. *Administrative Science Quarterly, 26,* 171–186.

Greene, J. G. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review, 12*(2), 91–116.

Heclo, H. (1978). Issue networks and the executive establishment. In A. King (Ed.), *The new American political system*. Washington, D.C.: American Enterprise Institute, 87–124.

House, E. (1993). *Professional evaluation: Social impact and political consequences*. Thousand Oaks CA: Sage.

Huberman, M. (1989). Predicting conceptual effects in research utilization: Looking with both eyes. *Knowledge in Society: The International Journal of Knowledge Transfer, 2*(3), 6–24.

Kingdon, J. (1995). *Agendas, alternatives and public policies*. NY: Harper Collins.

Leviton, L. & Boruch, R. F. (1983). Contributions of evaluation to education programs and policy. *Evaluation Review, 7*(5), 563–598.

Patton, M. Q. (1988). The evaluator's responsibility for utilization. *Evaluation Practice, 9*(2), 5–24.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text*, Thousand Oaks, CA: Sage.

Reimers, F. & McGinn, N. (1987). *Informed dialogue: Using research to shape education policy around the world*. Westport, CT: Praeger.

Rivlin, A. M. (1971). *Systematic thinking for social action*. Washington D.C. Brookings Institution.

Sabatier, P. A. & Jenkins-Smith, H. C. (1993). (Eds.), *Policy change and learning: An advocacy coalition approach*. Boulder CO: Westview.

Schon, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass.

Senge, P. M. *The fifth discipline: The art and practice of the learning organizations*. NY: Doubleday Currency.

Shadish, W. R. (1997). EVALTALK message. September.

Shulha, L. M. & Cousins, J. B. (1996). Recent developments in theory and research on evaluation utilization. Paper presented at American Evaluation Association meeting, Atlanta, November.

Shulha, L. M. & Cousins, J. B. (1997). Evaluation use: Theory, research and practice since 1986. *Evaluation Practice, 18*(3), 195–208.

Solo, P. (1997). Beyond theory: Civil society in action. *Brookings Review, 15*(4), 8.

Weiss, C. H. (1988a). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice, 9*(1), 5–19.

Weiss, C. H. (1988b). If program decisions hinged only on information. *Evaluation Practice, 9*(3), 87–92.

Weiss, C. H. (1980). Knowledge creep and decision accretion. *Knowledge: Creation, Diffusion, Utilization, 1*(3), 381–404.

Weiss, C. H. (1970). The politicization of evaluation research. *Journal of Social Issues, 26*(4), 57–68.

Weiss, C. H. (1967, April). Utilization of evaluation: Toward comparative study. In House of Representatives committee on government operations, *The use of social research in federal domestic programs*, Part III. Washington, D.C.: Government Printing Office, 426–32.

Wolfe, A. (1997). Is civil society obsolete? *Brookings Review, 15*(4), 9–12.