How Far Dare an Evaluator Go Toward Saving the World?[1]

Bob Stake

No two professional evaluators are the same but many use similar methods. Still, each will use a method in a somewhat idiosyncratic way. Especially in the interpretation of data, personality and experience have a play.

Professional evaluators come from many backgrounds. They have greatly different aspirations. As a group they are considerate people. They are ethical. They follow disciplined procedures to find the merit and worth of a program or other object. Oh, a few are rogues. And a few sell themselves to the highest bidder. But most of us evaluators are good people most of the time. We are specialists at recognizing differences among greater quality and less quality. We hope that our work will contribute to the making of a better world.

While sitting in a waiting room, I overheard a young woman say, "I'm an evaluator now." Answered by "Don't you do consulting any more?" "Oh, yes, but I get more attention if I call it 'evaluation.'" Some evaluators call their work "evaluation" because it gets more opportunity to help make changes for the better. In large and small ways, they hope to help save the world. Perhaps this is a bit of false advertising.

One evaluator I know is passionate about discovering perfidy, particularly bureaucratic deceit. Another I know seeks to balance evaluator voices with teacher voices. As for me, I find myself digging into issues of continuing professional education, regardless of the questions prioritized in the contract.

Advocacies. Most evaluators claim to make <u>dispassionate</u> searches for quality and dysfunction. They speak disdainfully of advocacy and promotion. Yet it is clear that many of us evaluators have strong feelings about certain matters and we promote those values in our work. Here are six advocacies common in evaluation studies:

---

[1] Paper delivered at EQRE, Pittsburgh, June 7, 2003.

1) We care about the evaluand, the object being evaluated. Often we believe in it. The *internal* evaluator is evaluating a part of his or her own organization. Barry MacDonald once said, "One should not evaluate a program if one does not support its goals." Seldom do we have a large conflict of interest, but often a *confluence* of interest. We hope to find the program working. We are disposed to see evidence of success.

2) We care about evaluation. We want to see others care about it. We want to encourage them to do it. We promote evaluation services, our own and those of our colleagues. We favor methods that evaluate well, and encourage others to use them too. It is an advocacy we flaunt.

3) We advocate rationality. We would like our clients and other stakeholders, our colleagues and heads of department to explicate and be logical and even-handed. We often pause in our data gathering or reporting to point out a way that the evaluand could have been run more rationally.

4) We care to be heard. We are troubled if our studies are not used. We feel evaluation is more useful if program participants take some ownership of the evaluation. Some of us, myself included, are strong advocates of self-study and action research. Even an external evaluation can profitably use input from stakeholders, including suggestions for design and interpretation. Some of us, not including myself, strongly support participatory evaluation in which certain stakeholders take responsibility for design, data gathering and resolving questions of merit and shortcoming.

5) We are distressed by under-privilege. We see gaps among privileged patrons and managers and staff and underprivileged participants and communities. We would organize some of the evaluation study to issues of privilege, conceptualizing issues that might illuminate or alleviate under-privilege, and assuring wider distribution of findings.

6) We are advocates of a democratic society. We see democracies depending on the exchange of good information, which our studies can provide. But also, we see democracies needing the exercise of public expression, dialogue, and collective action. Most evaluators try to create reports that stimulate action.

Although we are troubled by the possibility that our advocacies will cause us to search more vigorously for aspiration-based evidence than other, we cling to some advocacies more than to neutrality, believing these well-considered biases to be compatible with the interests of the profession, our clients, and society.

Each of us is more than an evaluator. We are complex human beings. Some of the things we do are part of our work and some are outside our work. We have political, spiritual, aesthetic and other advocacies. Some of the panorama of advocacy cannot help but become part of the evaluation study, even if we try to confine it to the rest of our life. Perceptions and values from any part of our lives may influence the interpretations we make.

Ethical constraints. It is an ethical responsibility for the evaluator to identify the affiliations and ideological commitments that might influence his or her interpretations, not only for the contractors but for the readers of reports, and of course, for the evaluators themselves. But there is no way for the evaluator to identify all predispositions, nor even *to know* them. We can expose ourselves a little, through vitas, biographical notes, previous reports, acknowledgements of preference and alliance, even indirectly in the ways we write, but an entire list of possible influences would be unhelpful, arbitrary, ever out-of-date, and ever incomplete. It is difficult to help others realize the biases, good and bad, to be found in evaluators.

The evaluators themselves sometimes see the conflict of roles: one the role of evaluator as finder and reporter of program quality, the other the preservation of quality and restorations to quality. A contract to discern quality *is not* a license to fix things. The client may want help fixing things, and may so specify in the contract, but there is some risk to that. Evaluators then are under some persuasion to over-attend to things they can fix and to neglect things they cannot. They may apply for the evaluation work mostly to get the remediation assignment. They may do good things for the program but leave it not thoroughly evaluated. The client may claim the program was evaluated when what really happened was some fixing. There are reasons for keeping the two tasks separate, evaluation and repair.

Generally we presume that successfully carrying out an evaluation study is a contribution to social well-being. Its report will describe its activity and problems and show, to some extent, the merit and shortcoming of the evaluand, its effectiveness, productivity, and impact. Some of that information should be useful to program managers, the staff, the service recipients and stakeholders in general and should add to the efforts of professionals, and reformers and others to make this a better world.

But those efforts, of course, are never enough to satisfy the great need for improvement. Sometimes the evaluator is encouraged to instruct the staff in the processes of evaluation. Sometimes the evaluator chooses to supervise the staff's gathering data. Sometimes the evaluator is persuaded by his or her discipline, fellow evaluators, or conscience to help the organization or all society to become a more rational or more democratic enterprise.

Evaluation addenda such as these are often little explored during negotiation of a contract. Some of them grow spontaneously during the evaluation work. Spontaneity and emerging sensitivity can be good but they sometimes violate evaluation ethics calling for full disclosure of purpose and practice.

Guiding Principles. I have studied documents of ethics and standards, looking for guidance on matters of advocacy and activism. Are both conflict and confluence of interest recognized as a problem? What do the standards say about advocacies? Best known statements of ethics for evaluators are two documents, the *Joint Committee Standards* (1994) and the *American Evaluation Association Guiding Principles* (www.eval.org).

For my review of the *Guiding Principles*, I enlisted some brief help from one of the authors, Will Shadish (see Shadish, Newman, Scheirer & Wye, 1995). He selected as relevant the same two principles I cite below, plus adding that Principles C1, C4. E3 and E5 suggest that "an evaluator should inform a client if there is reason to believe he or she might object to a particular value commitment." The most relevant two:

A2. Evaluators should explore with the client the shortcomings and strengths both of the various evaluation questions it might be productive to ask and the various approaches that might be used in answering these questions.

C3. Evaluators should seek to determine, and where appropriate be explicit about, their own, their clients' and other stakeholders' interests concerning the conduct and outcomes of an evaluation.

The Principles do not directly show concern about evaluator advocacy, nothing to draw attention to such as the six biases I mentioned earlier. Shadish pointed out that the drafters of *the Principles* were careful not to deal at the level of particular methodological approaches, seeking broad positions at a higher level of conception.

The *Joint Standards* were equally non-specific about evaluator value-orientations. Standard C3 calls for "full and frank disclosure", C4 calls for balanced reporting, and D11 calls for objective reporting. The document goes on to identify guidelines and pitfalls, getting more specific, but still not raising concern about confluence of interest, intent to remediate, or activism. There is implication that one should stick to the job of finding merit.

Let us look directly at two attractive advocacy-related approaches: participatory and democratic evaluation. Participatory evaluation (in independent external evaluation) is aimed at engaging the program staff in responsibilities normally belonging to an evaluation specialist, particularly: research design, data gathering, analysis and interpretation (Patton, 1996; Kemmis, 1986; & Fetterman, 1994). It is claimed that their involvement will make the study more relevant and increase the likelihood that the findings will be used for further program development. Sometimes program beneficiaries and other stakeholders also are invited to participate. Considering staff time needed, evaluation costs may be higher. Many of us expect that the quality of such an evaluation will be technically inferior and in some ways conceptually less deep. But the organization may be helped much more by participatory evaluation than by a conventional external evaluation. The participatory advocate claims that improving the organization can be worth more than the best possible determination of merit and worth.

Democratic evaluation methods are another powerful advocacy of some evaluation writers (MacDonald, 1977; House and Howe, 1999; Greene, 1996). This approach draws the attention of the evaluator to concerns of under-represented stakeholders, the people seldom heard, sometimes seeking ways of engaging them in dialogue about policy and program change. The evaluator may merely assure that some issues reflecting the values of the under-represented are included in the design or may, in the interpretation of findings, strive to make a persuasive case for their support. And it may make a special effort to see that favorable findings are made available to people who might contribute to that support.

Diversity. Whether few or many, efforts to refine the evaluand will diminish attention to the primary evaluation issues. Conceivably the evaluator would confine assistance to off-duty time, thus at no expense to the funding body, but that's not the way it's done. Whether by democratic evaluation or the exercise of any other advocacy, the evaluator expects it to be part and parcel of the evaluation work. And it is impossible for the evaluator's interpretation of evaluation findings not to be colored by all parts of his or her life and even its camaraderie, community and other affiliations.

Joint Standard A5 says:

> The evaluation report should describe the object being evaluated and its context, and the purposes, procedures, and findings of the evaluation, so that the audiences will readily understand what was done, why it was done, what information was obtained, what conclusions were drawn, and what recommendations were made.

It is implied that the evaluator has the descriptive capacity to tell the full story. It is implied that a clear description can be readily understood by the audiences. The reality is that, exercising all good faith, much will not be reported, and much will not be understood.

Furthermore, much will differ from evaluator to evaluator. We sometimes aspire to a professional practice by which--hypothetically--all evaluators evaluating a single evaluand would produce largely the same findings. But it is not an attainable aspiration, and to force it to happen would bring disaster. Evaluators cannot help but see things differently. Their findings will be different. Hopefully not often completely at odds, but that too will happen.

In the complex determination of program quality and accomplishment, there is no single reality we can capture. The only reality is that constructed by people, and people differ. Were we to agree completely on what we see, we would presume we are seeing correctly—and we never fully are.

We have an evaluation practice that is influenced by the value commitments of the evaluator and a set of operating standards that imply we can attain a widely-agreed-upon picture of merit and worth. Something needs to give. It could be that we should more effectively constrain our value commitments and search harder for meta-evaluation consensus, but we clearly should develop our standards and principles so that they deal better with the uncertainty and individuality of evaluating. Evaluators should be encouraged to “have a life” and to “have a dream” so their interpretations are enriched by their experience. Comprehensive, idiosyncratic interpretations are small steps toward saving the world.

References

Fetterman, D.M., 1994. Empowerment evaluation. *Evaluation Practice, 15,* 1, 1-15.

Greene, J. C., 1996. Qualitative evaluation and scientific citizenship: Reflections and refractrions. *Evaluation, 2, 277-289.*

House, E. R. and Howe, K. R., 1999. *Values in education and social research.* Sage.

Joint Committee on Standards for Educational Evaluation, 1994. *The program evaluation standards. How to assess evaluations of educational programs.* Sage.

Kemmis, S., 1986. *The action research planner.* Geelong: Deakin University Press.

MacDonald, B., 1977. A political classification of evaluation studies. In Hamilton, D., Jenkins, D., King, C, MacDonald, B., and Parlett, M., *Beyond the numbers game.* London: MacMillan.

Patton, M. Q., 1996. *Utilization-focused evaluation*. Sage.

Shadish, W. R., Newman, D.L., Scheirer, M. A., & Wye, C. (editors)(1995). *Guiding Principles for Evaluators.* New Directions for Program Evaluation, 66. Jossey Bass.