

# Is the Goal of Evaluation Accountability or Assistance?<sup>1</sup>

Linda Mabry  
Indiana University

Evaluators are typically obliged to, provide both formative reports to assist program personnel for purposes of program improvement and summative reports to funding agencies for purposes of accountability. The discrepancy between these two tasks frequently nudge working relationships with program personnel into adversarial relationships. Caught between the conflicting information needs and tasks of two immediate clients (program personnel and sponsors), evaluators may encounter internal political difficulties which spawn problems of accuracy in program representation and validity of findings. This paper will note the formative-summative, assistance-accountability tension both generally as a common professional dilemma and specifically as a problem of actual practice in the ongoing evaluation of the Multicultural Education Program for prospective teachers.<sup>2</sup>

## Blurred distinctions, purposes, and roles

The distinction between formative and summative evaluation is well-established in the literature of program evaluation<sup>3</sup> but, in practice, the distinction is less clear. Data gathered for formative evaluation by program personnel is also used for summative purposes when sponsoring agencies make decisions regarding continuation funding or early termination. In addition, formative evaluations in the early years of a grant-sponsored program are later synthesized into summative reports at the end of the funding period. In the intersection between these clients and their purposes, there is nearly inevitable tension encountered by evaluators.

Program personnel and their sponsors are, in one sense, partners in the endeavor of the program, but in another sense, their different roles and responsibilities can introduce an adversarial element into their working relationships. Program personnel expect ongoing information from evaluators to assist them in identifying and improving difficulties, but conveyance of the same information to their sponsors may put them at risk of funding cut-backs or termination. Funding agencies expect evaluators to deliver an external opinion as to the quality achieved in a program, often pursuant to a determination either to continue or to terminate the program-accountability decisions with high stakes for program personnel. Program personnel are typically interested in constructing a program which maximizes benefits to immediate stakeholders and use of locally available resources, while sponsors are often interested in developing a model program, a prototype transferable to other settings and independent of a particular environment<sup>4</sup>. The adversarial relationship which may exist or emerge between program personnel and sponsoring agencies may create enmity between program personnel and evaluators who must report to sponsors.

Moreover, the diverse, sometimes irreconcilable<sup>5</sup> information needs and rights of stakeholders beyond immediate clients complicate the evaluator's task, often intensifying

the obligation for summative information in a policy environment in which many social and educational programs compete for scarce fiscal resources.

The highly charged professional landscape shared by evaluators, program personnel, funding agencies, and other stakeholders not infrequently yields dual and sometimes conflicting responsibilities regarding assistance and accountability. Evaluators can simultaneously find themselves assuming or pressed toward such diverse roles as:

- detective determining program processes and outcomes
- archivist documenting program quality
- critical friend helping to identify strengths and weaknesses
- coach or consultant helping personnel to recognize and act upon possibilities for improvement
- mediator among contentious personnel or stakeholders
- advocate or public relations agent helping to protect the program from unfair or insensitive attack
- devil's advocate challenging staff to overcome blind spots
- judge of the extent to which goals were met or addressed
- cathartic sounding board for complaints and grievances
- spy after information the staff may prefer to keep from the evaluator
- critic when the client feels unjustly belittled by an evaluation
- idiot when the client feels misunderstood
- hanging judge when the client feels doomed by the evaluation report
- schizophrenic in trying to balance many tasks and stakeholder needs<sup>6</sup>

Roles are shaped by participants' understandings of the nature of knowledge and of truth and by their perceptions of ethical obligations incurred by those who are party to an evaluation. Consciously and unconsciously, they underpin behaviors and interactions by all parties. Some aspects of role are shaped by the evaluation approach or model<sup>7</sup> with its implications for epistemology underlying methodology. But any role clarity which a particular approach may offer an evaluator may merely complicate interactions, as clients and stakeholders are unlikely to be offered the same clarity. Moreover, an evaluator's strivings toward identification with one or another approach to evaluation may fall short of

the expectations of other professionals in the field.

In the quest for truth about a program, in the attempt to represent the program as accurately as possible, evaluators may please sponsors and satisfy their own senses of integrity. But evaluation loses an important *raison d'être* to the extent that it undermines support for particular programs rather than assisting them to attain useful goals.

### The Multicultural Education Program

An ongoing evaluation in its third year of appraising the externally sponsored Multicultural Education Program (MEP) will be briefly referenced as a practical illustration.<sup>8</sup> Prospective teachers enroll in MEP, taking courses at a state-sponsored university and completing practica in cultural immersion settings. The program began in 1997, when it was funded for three years. The funding awarded to the program was drastically lower than the budget proposed, with the result that many program activities and functions have been forced to operate at minimal and subminimal levels through no fault of the program staff.

In this case, as in many program evaluations, funders expect annual evaluation reports. It is readily apparent that early reports summative in character would be premature, that the development of a successful program is unlikely in less time than it takes to prepare a teacher for the classroom. Yet, the formative reports of years 1 and 2, which have helped program personnel identify and address problems, problems which may be elegantly resolved in time, have seemed punitive to program personnel. Increasingly, project administrators have behaved defensively in interactions with evaluation personnel, even to the point of restricting access to some data-creating not only internal political difficulties but also dilemmas regarding accountability and validity.<sup>9</sup>

Members of the evaluation team increasingly encountered blocked access to data as program staff moved to protect their efforts and reputations, their distrust growing. The evaluators responded with surprise—we were there to help, after all, to tell the truth and set them free—and redoubled efforts at sensitivity and persuasiveness. But, despite the ease of the original agreement regarding the evaluation design, the program's staff early enthusiasm about our partnership on behalf of the program, there was no mistaking our stripes: We were there to try to find out everything and to write it up, even the unflattering parts, and our clients knew it and winced. Negotiating the final draft proved a precarious and

lengthy process which similarly left all parties bruised. Members of the evaluation team are still wondering: But what were we supposed to do? Is evaluation about scientific rigor, or providing service to our comrades-in-arms, or contributing to a common good beyond the program which it too supports?

### **Dilemma**

One problem here is a difficulty with priorities. Evaluation loses an important *raison d'être* to the extent that it undermines support for particular programs rather than assisting

them to attain publicly useful goals. If the evaluator's first allegiance is to the program and if the program is managing to struggle to its feet promisingly, then the program must be assisted on the one hand while it is simultaneously protected on the other. In the case of the MEP, this has involved softening the data in presentation, qualifying negative interpretations, and taking care to attribute difficulties carefully so that it does not appear the program is responsible for the practical implications of the draconian budget cuts imposed on them or for the evaluation forced upon them before they are ready to present outcomes.

But evaluation loses its claim to accuracy and credibility when hard facts are muted, whatever the good (or not so good) cause or intent. For the profession to maintain its opportunity to contribute to programs and to the social good, it may also need to maintain its reputation for cold-eyed observation and clear-headed analysis. As with research, this reputation is already under challenge from charges of positive bias (evaluators protecting their livelihood and opportunity for practice by pleasing clients who may recommend or rehire them) and charges of political naivete (evaluators or their reports manipulated by clever stakeholders).

The dilemma exists somewhere in the frayed overlap between the modern world and its optimism regarding the possibility of attaining and representing truth and of discovering and progressing toward something good or better and the postmodern world and its pessimism regarding the impenetrably personalistic nature of truth and of its insistence that anything good for one group oppresses another.<sup>10</sup> As postmodern ambiguity may imply, given a charitable spin, it may be that it is more productive for us to worry and wonder than to turn our attention too quickly, too comfortably elsewhere. Opportunities for evaluators to consider issues both in the abstract and in the particulars of ongoing practice facilitate the development of sensitivity to role dynamics and multi-faceted responsibilities, facilitate the development of expanded repertoires of response to difficulties of practice, and facilitate the development of collegial networks for professional advice and consultation.

<sup>1</sup> Paper presentation to the Septième Colloque Annuel de la Société Québécoise d'Évaluation de Programme, Saint-Hyacinthe, Québec, Canada

<sup>2</sup> a pseudonym

<sup>3</sup> Scriven, 1990, pp. 19-64; 1991; Worthen, Sanders & Fitzpatrick, 1997

4

DeStefano, 1992

<sup>5</sup> Mabry, 1997a

<sup>6</sup> For fuller discussion, see Mabry, Christina & Baik, 1998

<sup>7</sup> See, for example, Madaus, Scriven & Stufflebeam, 1987; Shadish, Cook & Levitton, 1991; Worthen, Sanders & Fitzpatrick, 1997.

<sup>8</sup> Mabry, 1997b; Mabry & Baik, 1998. Another illustration of this type of difficulty in the program evaluation literature may be found in Walker, 1997

<sup>9</sup> Both descriptive and evaluative validity are imperiled by inaccess to data, as described in Maxwell, 1992

<sup>10</sup> See Mabry, 1997a; Roseneau, 1992; Wakefield, 1990

## References

DeStefano, L. (1992). Evaluating effectiveness: A comparison of federal expectations and local capabilities for evaluation among federally funded model demonstration programs. Educational Evaluation and Policy Analysis, 14 (2), 157-168.

Mabry, L. (1997a). A postmodern text on postmodernism? In L. Mabry (Ed.), Advances in program evaluation: Vol. 3 Evaluation and the Post-modern dilemma (pp. 1-19). Greenwich, CT: JAI Press.

Mabry, L. (1997b). Year 1 evaluation report: Inter-Campus Enhancement of Language Minority Teacher Recruitment and Bilingual Education program. Bloomington, IN: Indiana University.

Mabry, L. & Baik, S. (1998). Year 2 evaluation report: Inter-Campus Enhancement of Language Minority Teacher Recruitment and Bilingual Education program. Bloomington, IN: Indiana University.

Mabry, L., Christina, R., & Baik, S. (1998). Impact of conflicting obligations on role, responsibility, and data quality. Paper presented at the annual meeting of the American Evaluation Association, Montreal, Canada.

Madaus, G. F., Scriven, M. S. & Stufflebeam, D. L. (Eds.). (1987).

Evaluation models: Viewpoints on educational and human services evaluation. Boston: Kluwer-Nijhoff.

Maxwell, J.A. (1992). Understanding and validity in qualitative research. Harvard Educational Review, 62 (3), 279-300.

Rosenau, P. R. (1992). Postmodernism and the social sciences: Insights, inroads, and intrusions. Princeton, NJ: Princeton University Press.

Scriven, M. (1990). Beyond formative and summative. In M. McLaughlin & D. Phillips (Eds.), Evaluation and education at quarter century (pp. 19-64). Chicago: University of Chicago/National Society for the Study of Education.

Scriven, M. (1991). Evaluation thesaurus (4th ed.). Newbury Park, CA: Sage.

Shadish, W. R., Jr., Cook, T. D. & Leviton, L. C. (1991). Foundations of program evaluation: Theories of practice. Newbury Park, CA: Sage.

Wakefield, N. (1990). Postmodernism: The twilight of the real. London: Pluto Press.

Walker, D. (1997). Why won't they listen? Reflections of a formative evaluator. In L. Mabry (Ed.), Advances in program evaluation: Vol. 3 Evaluation and the post-modern dilemma (pp. 121-137). Greenwich, CT: JAI Press.

Worthen, B. R., Sanders, J. R. & Fitzpatrick, J. L. (1997). Program evaluation: Alternative approaches and practical guidelines (2nd ed.). New York: Longman.