

THE SOCIOLOGY OF PROGRAM EVALUATION

Dan S. Green, Pennsylvania State University

INTRODUCTION

As witnessed by the last few years, evaluation, program evaluation or evaluation research has evolved from a relatively obscure phenomenon into a commonly-known process. Although the process of evaluation is probably nearly as old as the phenomenon of formal organization, our referent is the formal evaluative process of social programs, based on scientific procedures. Program evaluation has also become, or is rapidly becoming, a major sub-field in the discipline of sociology. It would have been unheard of, for example, not too many years ago, to have a section on program evaluation at a professional sociological meeting; it is now rather commonplace.

Program evaluation has achieved the status in sociology that it has had for some time in education and psychology. It must not be forgotten, however, that program evaluation is a method or set of procedures in the same sense as scientific method, hence it transcends all of the social sciences and education, and is not the property of any given discipline (Oetting, 1976). Although the technology and procedures of evaluation have been altered to fit the requirements, biases, and needs of various disciplines, its base is still, and must be rigorous scientific procedures. This commonality will rightfully keep evaluation from becoming the property of any discipline or from becoming a separate discipline.

The diffusion and acceptance of program evaluation in sociology, as well as other disciplines, seems to be a function of the growing emphasis toward applied endeavors. Program evaluation seems to have greatly served to legitimate applied sociology.

THE SOCIOLOGY OF EVALUATION

There are several issues that could be discussed under the rubric of the sociology of evaluation. Among these are why evaluations of social programs are not performed and why evaluations are not routinely called for in policy formation, the types of evaluation, in house versus outside evaluation, the status and role of the evaluator, evaluation nomenclature, role conflicts in evaluation, the politics of evaluation findings, the purpose of an evaluation, the

training of evaluators, the status of evaluation as a research endeavor, and the use and abuse of evaluation data.

The sheer complexity of the evaluation process, the plethora of problems surrounding the process and the significance of evaluative data together require, or perhaps, demand, detailed sociological analysis. Any such analysis may be performed on two levels. One level would be an analysis in terms of pure research--that is, the sociological analysis of some aspect of the evaluative process geared to validate a hypothesis or sociological pattern. A second level, more appealing to us, is applied analysis directed toward using the traditional sociological perspective to attempt to understand and perhaps unravel some of the problems and concerns of program evaluation. The point of view from which I shall argue is that of a student and teacher of evaluation, and perhaps more significantly, that of three years experience as an in-house program evaluator with a Project Follow Through model. Project Follow Through, a compensatory educational program consisting of over a dozen educational models for poor children, has been acclaimed as the largest social experiment ever launched, and the most extensively evaluated social program in the history of man (Maccoby & Zellner, 1970).

The Status of Evaluation

According to Weis (1979:9), "Evaluation is sometimes regarded as a lower order of research, particularly in academic circles, than 'basic' or 'pure' research." She goes on to note that evaluators are considered the drones of the research fraternity, but in reality, "program evaluation calls for a higher level of skills than research that is more under the researcher's control." Clearly, program evaluation as a research endeavor in sociology has not been accorded a very high status and may have been considered more as a technical effort than as legitimate research. The lower status of evaluation *vis a vis* pure or basic research seems to have been related to two factors. One is that program evaluation was not a sociological endeavor, that is, as defined by the discipline.

Secondly, as an applied endeavor, it was accorded the lowly status given to applied research. A related issue was that evaluation was directly related to social engineering which has not always had a high degree of respectability within the discipline.

The status of program evaluation in sociology has clearly risen since Weiss' 1972 statement, as indicated by the appearance of formal courses, textbooks, special summer programs, journal articles, positions in sociology departments, and sections and presentations at professional meetings. However, its status relative to basic research is still less clear. Its status is linked to that of applied sociology which is on the rise. Program evaluation will continue and will thrive, however, regardless of its status within the discipline. However the opportunities for training within sociology, are related to its status and that is a concern. The training of evaluators within sociology should be thoroughly studied in terms of focus of the training in related disciplines. One concern that arises in this context is that it seems that few sociologists are actually involved in informal evaluations. Generally, they seem to have been more involved in formal evaluations. Generally, they also seem to have been more involved in studying and reflecting upon evaluation. This reflects a bias against applied sociology and also has many implications for the type of training that is being provided in sociological programs.

THE SCOPE OF EVALUATION

The scope of any evaluative effort is a subject of some controversy. Some would argue that the ideal evaluation is a comprehensive evaluation. For example, Rossi, Freeman, and Wright (1979:45) write that a "comprehensive evaluation refers to studies that include monitoring, impact, and ex post facto cost-benefit or cost-effective analyses." Ideally, they state, all three will be a part of all evaluations. Yet Weiss (1972:15) claims that the comprehensive or "all-purpose" evaluation is a myth.

Experience working in a social program soon teaches one the difference between the ideal and what is practical, feasible, or permissible in terms of finances, time, effort, and technological capacity. In terms of an ideal evaluation, some sociologists also argue for some type of process or system evaluation. As

important as evaluation is, we cannot forget that the goal of a program is to serve people and there is never enough money or resources to adequately accomplish that goal. Diverting dollars to evaluation, although necessary, must be tempered.

Project Follow Through has been evaluated over a decade by the Department of Education, Abt Associates, SRI, Huron Institute, Nero and Associates, and others as well as by continuous in-house evaluations. Literally, millions of dollars have been spent evaluating the program and as yet the "ideal" evaluation has not been achieved. For example, the in-house evaluation (Green, *et al.*, 1977) that were routinely performed by our Follow Through model, Behavior Analysis Follow Through (BAFT) considered:

1. Daily and weekly monitoring of student performance.
2. Annual achievement testing of all students.
3. Weekly and monthly monitoring of the classrooms by on-site personnel and training consultants.
4. Annual consumer satisfaction surveys.

There were neither ideal nor comprehensive, yet, they provided us with the information that we decided was needed to reasonably evaluate the effectiveness of the program. The cost of these efforts, both in terms of dollars and resources, was staggering.

On three separate occasions, when presenting BAFT program results at professional meetings, I was asked if we had performed any process or system evaluation to see how and why the program worked. My response was simply that we knew, at least to our satisfaction, that the program worked; we were not terribly concerned with why.

TYPES OF EVALUATION

Given that the ideal evaluation is usually impossible, or at least impractical or unfeasible, then one must determine which type of evaluation to use. Generally, there are four basic types of evaluation: *outcome* (goal-oriented), *process* (evaluating what goes on within the program), *consumer satisfaction* (a measure of the satisfaction of the program consumers with the services), and *cost-effectiveness* (a measure of the cost as related to the service).

There seems to be some sort of discipline

bias at work. Sociologically trained evaluators seem to prefer process or structural evaluations; this would seem to be related to our affinity to the importance of various functions or the structure of the organization as explanatory variables. Sociologists shy away from goal-oriented evaluation which somehow focus upon the individuals served by the program. Psychologists, are more concerned with goal-oriented evaluation. Also, the preferred type of research design relates to ones discipline. Evaluators frequently support such biases without being aware of them. Since we argue that evaluation cuts across all the social scientific disciplines, such biases or limitations may unconsciously dictate the type of evaluation one uses.

Consumer satisfaction, sometimes referred to as social validation, is a relatively new type of evaluation procedure, and is noted here because of this, and because, as a variation of survey research, it should be particularly appealing to sociologists (Jackson, Green, & Fixsen 1976). In the BAFT project, we found consumer satisfaction surveys to be most beneficial. They seemed to have tremendous public relations value, provided us with a measure of social validity, provided several unique sets of consumers with the opportunity to give us both positive and negative feedback, and pointed out weak points in certain program functions which could then be corrected. Few social scientists, with the exception of economists, engage in cost-benefit analyses and the paucity of their use is certainly no indication of their value. Generally, there are two drawbacks to cost-benefit analyses: they are exceedingly complex, and most program managers know their programs are expensive--maybe too expensive--and prefer not to have this fact systematically and empirically documented.

Since cost-benefit analyses are so rarely performed, and because of the political pressure not to perform them, they will be given a low priority. Consumer satisfaction surveys are easy, fun, relatively inexpensive, and produce a nice set of data that is easily analyzed. Some problems arise, however, in that too frequently consumers are satisfied when they should not be, or, consumers are too easily satisfied by an intervention or attention and tend to rate any program, when, in fact, the goals, are not being accomplished. We all

know, for example, that it is quite easy to keep students happy or pleased in a classroom situation--whether or not they are learning becomes a moot point, but they are happy. Similarly, many studies of compensatory educational programs have discovered that parents are pleased with the program regardless of type of program, duration, or the actual results (McDill, McDill, & Sprehe, 1969:43-44). Campbell (1969:426) refers to this as the "grateful testimonial."

Process evaluation, measures what goes on within the program. I believe that any analysis of the program organization or functions is clearly much less significant than outcome or goal attainment measures. Outside evaluators at times fail to see that process evaluation is a luxury that few programs can afford. This type of evaluation is more clearly in the realm of sociology where such factors as organizational analysis, and manifest and latent functions, are considered important. To those working in a program, that which is most important is the outcome. Does the program work?

That leaves us with goal-oriented or outcome measures as the most preferred and most pressing type of evaluation. It is critical to determine, as soon as possible, if the program goals are being met. If the outcome is favorable, then one may decide to perform another type of evaluation. If not, there is little need for any other type of evaluation. An exception here would perhaps be some type of process evaluation to see why the program is found to be ineffective; this, however, is costly, and often, when a program is found to be ineffective, the problem is with the goals, which are unclear or difficult to measure, or with the personnel charged with implementing the goals--hence the problem is rather obvious. The BAFT program circumvented this latter issue by obtaining weekly data regarding project implementation. These objective data were supplemented with periodic in-house daily or weekly and outside monthly observations of the teachers and the classrooms.

Internal vs. External Evaluation

Another issue that has been the subject of much debate is in-house (internal) vs. outside (external) evaluation. According to Weiss (1972), this issue revolves around the factors

of administrative confidence, objectivity, knowledge of the program, potential for utilization of results, and autonomy. The first factor, administrative confidence, is, in part, a status problem; program administrators may have more competence than their staff members. On the other hand, administrators may question the status of an outsider, viewing them as too remote, or too abstract to provide any useful or practical information.

Regarding objectivity, it is frequently stated that insiders will be less objective than outsiders. There are potential problems which may make this so, such as an inside evaluator having to tell his boss that the program is not working, or the program manager to subtly inform the inside evaluator of the consequences of a poor evaluation. But with honest people, there is simply no real threat of loss of objectivity with an inside evaluation--at least no more so than with an outside evaluator. The question of objectivity boils down to the integrity of the evaluator--insider or outsider. An outside evaluator, or a contractual evaluation team, may or readily be less than objective in order to appease the program management. Who would want to maintain a evaluation contract with someone who comes forth with a negative evaluation?

Insiders appear to have a significant edge with regard to knowledge of the program. Although outsiders can learn about the program, the time spent learning can be a costly process. Regarding utilizing the results, the status of an outside evaluator seems to allow him/her to make the necessary recommendations and induce the managers to heed the results. But again, this is, in part, a matter of integrity. Nevertheless, the higher the status of the evaluator generally the more attention the results will receive. Considering autonomy, the outsider will generally have the edge here because of the nature of his/her status to the program.

In our Follow Through program, we had the best of both worlds; that is, we routinely performed our own in-house evaluations, and outside evaluations were also routinely being performed. As in-house evaluators, we did not feel that objectivity was a problem--at least no more so than in any complex research endeavor.

PSEUDO EVALUATION

Related to the above discussion is the purpose and nature of the evaluation. Too frequently, an evaluation or something denoted as an evaluation--usually some hastily concocted data--is performed to legitimize the program, the staff, or management, or to satisfy the terms of a grant. This type of "evaluation" can destroy the entire concept of evaluation. I shall denote this as "evaluation-type data used to legitimate a program." This type of pseudo-evaluation can be contrasted with any formal evaluation used as a useful, legitimate tools. Suchman (1972:81) refers to such pseudo-evaluations as "eye-washes": selecting only those aspects that make the program "look good" on the surface; "white-washes": the avoidance of any objective appraisal; and, "the posture": an attempt to use evaluation as a "gesture" of objectivity.

EVALUATION TERMINOLOGY

A final area that we wish to discuss is the language of evaluation. Those who are familiar with the field of evaluation realize that it is beset with the same terminological and conceptual confusion that exists in all of the social sciences. But this should be no surprise since program evaluation, is in large part, a product of the social sciences. For example, we seem to be uncertain about the basic nomenclature of the evaluative process. Some try to delimit between the terms *evaluation*, *program evaluation*, and *evaluation reserach* (Franklin & Thrasher, 1976 24; Suchman, 1967). The terminology becomes much more complex and beclouded when one turns to the types of evaluation. Steele (1973), cataloging the various types, presents thirty-seven types arranged in six major groups.

There are several legitimate types of evaluation, each of which serves a specific function geared to provide one with information about the program or the outcome of the program, and it simply doesn't make any difference whether we call the process evaluation, program evaluation, or evaluation research. The variety of types present enough problems in themselves, however, most active evaluators are able to perform without worrying about whether a process or structural evaluation is really the same thing.

Rather than worrying about the terminology, we should be concerned about the difference between valid and invalid evaluations. Good evaluations are grounded in accepted scientific procedure; and bad evaluations are questionable in terms of the scientific procedures used.

EVALUATION "TRUTHS"

I would like to conclude with some unfortunate "truths" of the evaluation process which if not confronted by rigorous sociological analysis, may make all that I have said wholly academic. These "unpleasant truths" are paraphrased from Zusman and Bissonette (1973:122-3).

1. Large-scale, sophisticated, systematic evaluation of most programs is too complex and expensive to be a reality.
2. Many evaluations are poorly designed, underfunded, and forced on programs unwilling or unable to use the results.
3. Many evaluations are likely to be rejected, ignored, misinterpreted, or simply unnoticed.
4. Evaluation results, when accepted, are likely to be distorted when made operational.
5. The purpose, design, conduct, and use of evaluations are always subject to evaluational and political bias.
6. Resistance to evaluation will be ever present and multifaceted before, during, and after effort.
7. Given that all obstacles to evaluation are removed, the availability of evaluative data will still lag behind the exigencies of decisionmaking.

To counter these "truths," they offer five "realistic guidelines" for evaluators:

1. Halt the mounting fervor and crusade regarding the benefits of and the need to evaluate every program.
2. Evaluate programs early, before vested interests have solidified and organizational inertia has set in.
3. Conduct no evaluation at all where there is not strong likelihood that it can be done with scientific accuracy.
4. Widely publicize results and implications of an evaluation in concrete and simple language directly related to program procedures.
5. Report findings in clear, simple language

designed for program administrators rather than researchers.

The Zusman and Bissonette claim that evaluations of many kinds of complex human service programs cannot be done at all but, I think that any program can be evaluated and all should be. Some programs, perhaps relatively simple ones, about which there are no questions, may need no evaluation--but still how does one really know the program works in the absence of empirical data? Some programs cannot be evaluated because of lack of funds or expertise--but this says nothing about whether they should be evaluated. Programs without clear goals or orientations would seem to be those sorely in need of evaluating to point out the unclear goals or lack of orientation.

I would argue strongly that all programs should have a set of clearly articulated, objective, measurable goals. Given these, a clear orientation will necessarily follow. It is imperative to point out programs without clearly articulated goals. Evaluation can do this, and the evaluator could assist in clarifying the goals. Too frequently programs are established and implemented with little thought about goals or how to measure goal attainment. This is a clear case of putting the cart before the horse. We must be much more cognizant of program design and work toward teaching programmers to design their program around clearly articulated, measurable goals. In our Follow Through Program, for example, the instructional program was designed around a single objective goal: and one one-half years of academic progress for each year of school as determined by a year-end achievement test. The crucial point is that the goal preceded the design of the program. Given this, evaluation becomes an intrinsic part of the program design. Following this strategy would also compel us to give as much attention to program design as we have to evaluation. Here we have much to learn from our colleagues in business who have their MBO Management By Objectives (MBO) philosophy, and in education with their learning objectives and competency-based learning. Evaluation is an important and socially significant endeavor. There are many issues that are part of or closely related to the evaluation process that require further elucidation and analysis; the unique perspective of sociology is aptly geared for this task.

REFERENCES

- Campbell, Donald, 1969. Reforms as experiments. *American Psychologist* 24: 409-29.
- Franklin, J.L., Jean H. Thrasher, 1976. *An Introduction to Program Evaluation*. New York: Wiley.
- Green, Dan S., Darwin J. Easkins, John W. Scott, and Lynn c. Weis, 1977. "The Behavior Analysis Follow Through" evaluation strategy: A multifacted approach. *Evaluation* 4: 78-80.
- Jackson, Don A., Dan S. Green, Dean L. Fixsen, 1976. "The social validation of applied behavioral programs." Paper American Psychological Association Meeting.
- Maccoby, Eleanor M. Zellner, 1970. *Experiments in Primary Education: Aspects of Project Follow Through*. New York: Harcourt, Brace, Jovanovich.
- McDill, Edward, M.S. McDill, J.T. Sprehe, 1969. *Strategies for Success in Compensatory Education*. Baltimore: Johns Hopkins Press.
- Oetting, E.R., 1976. Evaluative reserach and orthodox science. *Personnel & Guidance J* 55 11-15.
- Patton, Michael Q., 1978. *Utilized-Focused Evaluation*. Beverly Hills, CA: Sage.
- Rossi, Peter H., Howard E. Freeman, Sonia R. Wright, 1979. *Evaluation: A Systematic Approach*. Beverly Hills, CA: Sage.
- Steele, Sara M., 1973. *Contemporary Approaches to Program Evaluation: Implications for Evaluating Programs for Disadvantaged Adults*. Syracuse, NY: Eric.
- Suchman, Edward A., 1967. *Evaluative Reserach*. New York: Sage.
- Suchman, Edward A., 1972. Action for what? A critique evaluative reserach. Pp. 52-54 in Carol H. Weiss ed., *Evaluating Action Programs*. Boston: Allyn Bacon.
- Weiss, Carol H., 1972. *Evaluative Research: Methods for Assessing Program Effectiveness*. Englewood Cliffs, NJ: Prentice-Hall.
- Zusman, Jack, Raymond Bissonette, 1973. The case against evaluation (with some suggestionns for improvement). *International J of Mental Health* 2 111-125.

BECNEL & LEVY From Page 22

Hawley, R.W., 1979. Marriage encounter experience and its effects on self-perception, mate-perception, and marital adjustments. Unpublished doctoral thesis, U. of Nebraska-Lincoln, 1979.

Kephart, W.M., 1972. *The family, society, and the individual*. 3rd Ed., New York: Houghton Mifflin.

Mace, D., 1979. Marriage and family enrichment - a new field? *Family Coordinator*, 1979, 287, 409-419.

Milholland, T.A., 1975. The effects of marriage encounter on self-disclosure, trust and marital satisfaction. Unpublished doctoral Thesis, Texas Tech., 1979.

Powell, John, 1974. *The secret of staying in love*. Niles, Illinois: Argus Communications, 1974.

Otto, H.A., 1975. Marriage and family enrichment programs in North America. Report and analysis. *Family coordinator*, 1975, 24, 137-143.

Ragula, R.R., 1975. Marriage encounter: What makes it work? *The Family Coordinator*, 1975, 24, 153-159.

Smith, R.M., Scott, J.P., Shoffner, S.M., 1976. Marriage and family enrichment: A position-paper. University of North Carolina at Greensboro, 1976.

Stinnett, Nick, Montgomery, James E., and Collins, Janet. Marital needs satisfaction of older husbands and wives. *J. of Marriage & Family*, 1970, 428-434.

Van denBos, Gary R. and Wolf, Lonn A., 1970. *Experiential focusing: New research tools*. Unpublished, 1970.