The relationship between social program evaluation and social policy and program development

Roger Lind

Risken är stor att utvärdering av social utveckling och sociala program omedvetet påverkas bl. a. av politiska och kommersiella förväntningar, hävdar Roger Lind, som har erfarenhet främst av amerikanska förhållanden.
   Roger Lind är professor i social work, Louisville, USA.

A number of years ago the British social policy thinker Richard Titmuss suggested in an article in an American journal that "criteria of moral progress are found in the growing power of altruism over egoism". (15,5) That is not a formulation which is particularly well-received by an American audience, even, I regret to say, in social work, which is of course attempting to be responsive to its surroundings and to appear as hard-nosed and specifics-oriented as the rest of the system.

Be that as it may, I suggest that a social program is really a means for giving form to the expression of human concern; even all the ideological differences, political jockeying, psychological aberrations and material self-orientations are features of human life which set the limits on the expression of human concern rather than negate it.

The trouble with the above formulation of social programs is that, taken at its simplest, it allows full play to cupidity, egomania and such warped manifestations of the inhumanness, or lack of civilization, among people and is therefore seen as a misstatement of the situation. But part of the problem is to keep a focus on the purpose, and to structure or design the forms to the greatest possible degree to accord with the purpose. Evaluation can then be seen as one of the many tools used to serve that end.

We desire information regarding social programs in order to increase our ability to perceive, interpret and evaluate results. To do that we try to make the methodology of evaluation more scientific, more technically exact, and thus to decrease the likelihood of disagreement, or differing interpretations of the same behaviors or events. This often results in our losing sight of the purposes we have in developing social programs. This goal-displacement is similar to the fact that in our efforts to ensure agreement and reduce criticism of programs we forget what it is we are really trying to do.

Three levels for social program evaluation

In their recent book, *An Introduction to Program Evaluation,* Franklin and Thrasher state that "... program evalution is defined as the determination and assessment of the results (outcomes/impacts) of program activities." (6, 23) Social program evaluation can be viewed on at least three levels. The first or immediate level would be that of evaluation as a detailed delineation of the results of a program. The second, intermediate, would be as an instrument for program improvement, adjustment or development. The third or ultimate level would be program evaluation as a means of assessing the extent to which the program fulfills its policy purpose (s).

The connection between social program evaluation and social program and policy development is that of the chain of logic, according to which nested elements relate to one another. Most current efforts regarding program evaluation are concerned with strengthening technical aspects in order to increase explanatory power and consumer confidence in evaluation as an instrument for decision making. (1; 3;10;12) The literature of applied social research is filled with references to studies gathering dust. This suggests that at the present time inadequate use is being made of program evaluation as an instrument for program development. This also suggests that it will be some time, at least in the United States, before we should expect to see much relationship between program evaluation and the development of social policy.

The role of evaluation in the United States.
The role of evaluation differs greatly in the United States and Sweden. This is due to a number of factors, among which I consider a few to be of special significance. The splintering of governmental jurisdictions in the United States is such a factor, as is the proliferation of social programs brought about by this splintering and by the relative autonomy of voluntary agencies. Both of these factors are reinforced by the ideological importance of pluralism in the United States.

Following from these, there is a lack of comprehensive program sponsorship, and a low degree of consensus on program goals and methods.

The role of elected public officials in regard to the operation of and surveillance over social programs is much less closely involved in the United States than in Sweden, and the officials are less likely to be knowledgeable about the programs.

The primacy of business values of economy and efficiency in American thinking plays an important role in determining what kinds of measurements are considered appropriate and valid, as to the emphasis on technology as a key to problem solving, and the tendency to think of individuals as instrumental rather than as goal values.

My remarks are based primarily on the American scene, with which I have greater familiarity, and so

while they have application at many points to issues of program evaluation in Sweden the applicability should not be assumed to be on the order of one to one.

If program evaluation is to have policy relevance it clearly must be linked up in its design with the policy purpose. Under the conditions, outlined above, in the United States, "program" and "agency" or "organization" are virtually never coterminous. It is therefore impossible to restrict one's attention to a single agency in doing a program evaluation if the intent is to assess the extent of fulfillment of policy purpose by means of program implementation. To illustrate, an evaluation of the child abuse program within a specified geographical area cannot be accomplished by studying the efforts of a particular agency on behalf of abused children, unless that agency is charged with responsibility for care, treatment and prevention on behalf of all children in the area.

A wide range of definitions for the term "program evaluation"
The literature on program evaluation offers a wide range of definitions for the term, some narrowly couched and calling for close adherence to strict requirements of scientific limits, others much more broadly construed and showing very little concern with methodology of proof. Two factors lead to the emphasis on a broad meaning to the term. One is the present level of development of evaluation techniques, as this limits our ability to solve the problems which seriously limit the generalizability of a specific study. (5; 18) The other is the always-existing need to do our best in improving the specific program under study.

For those who try to work on both aspects of this problem, immediate program improvement and generalizability, what limits the extent of this effort is the availability of resources. In the face of resource limitations the aspect which yields is generally the more expensive one, which is that of strict adherence to scientific requirements.

Franklin and Thrasher, cited above, argue that generalization is not a concern of evaluation research; the focus is simply on the program under

investigation. That is, they see evaluation research as always initiated "... by the felt need of managers and policy makers." (6, 38) This appears to me to be an inaccurate description of the role of such efforts.

In order to be able to work on both aspects of the problem, immediate use and generalizability, it seems necessary to approach the process of program evaluation with attention to both outcomes and processes. We need to be able to speak with a degree of knowledge about both results and effort. The question of whether the program has been applied to the problem, whether the population has received the service, must be responded to, and only then is it possible to know whether the results, outcomes, manifestations which we observe and report are, indeed, the consequences of the program intervention. It is for these reasons that I find it necessary to be concerned with both process evaluation and outcome evaluation, and thus to consider an evaluation which does not pay attention to both to be incomplete, however sophisticated.

Are we fooling ourselves?

A central issue in research is, "Can we be sure we're not fooling ourselves?" This question is central to all our interventions in social life, and the scientific method with its prescriptions regarding research design provides us with some degree of reassurance. There are stringent requirements for sampling, measurement, statistical tests, significance levels, experimental design, and so on. Studies which do not attempt to work on those issues are essentially intended to help do a single, specific job better, not to enable us to *generalize* to other situations with a reasonable degree of confidence.

But experience tells us that when one has developed an approach which works in one set of circumstances, one has a tendency to want to use it in others as well. The result is that generalizations are made even when that was not the intent.

A force which supports and encourages this tendency is the interest of politicians and the lay public in the solution of social problems and the successful implementation of social programs. The push for solution leads to pressure for expanding the use of what appear to be successful approaches, and lack of acquaintance with or commitment to the cautions of scientific method can mean impatience with its requirements. In the United States at least, the nature of electoral politics favors quick expansion of demonstration projects which appear to work. A recent example of this was development of the community action approach in the War on Poverty in 1965, (9) growing out of brief experience in the demonstration projects initiated under the Juvenile Delinquency and Youth Offenses Control Act of 1961. (8)

The point to be made is that it is unrealistic to assume that attempts to apply successful approaches will be limited to the project in which findings of success were made. Such an idea serves more as a convenient rationale to disregard some of the requirements of careful study than as a reasonable description of the narrow limits within which findings will be applied.

Early emphasis in the development of evaluation of social programs concentrated on the improvement of technical aspects of design and measurement. The low rate of utilization of results has been a factor in shifting attention to questions of feedback and the politics of evaluation. The latter subject has generally been concerned in part with relationships between and among program administrators, practitioners and evaluators. The concern involves underlying differences in orientation and features of the interactions which are friction-promoting or friction-reducing. (17, 98–107)

The role of political factors

Concern for the political context within which program evaluation occurs draws our attention to the *necessity* of recognizing and assessing the role of political factors such as negotiations, trade-offs and power allocations in order both to complete a successful study and to increase the likelihood of utilization of findings of success. (1, 172-9, 496–506)

The emphasis in the United States on program evaluation as an administrative tool for the use of program managers gives it more the character of an aid in program planning than as a component of long-range policy development. This emphasis may

well be a result of infrequent utilization of managers and evaluators for policy development. The American doctrine of "separation of powers" in government may play a role in this, as does the splintering of governmental jurisdictions and of program administration mentioned earlier.

One consequence of the focus on the program manager as both initiator and recipient of program evaluation studies is that evaluation runs the risk of being used by dominant coalitions in the organization, and the evaluator of becoming simply the tool of the managers. Powerful vectors come together in program evaluation:

Figure. Influences in evaluation.
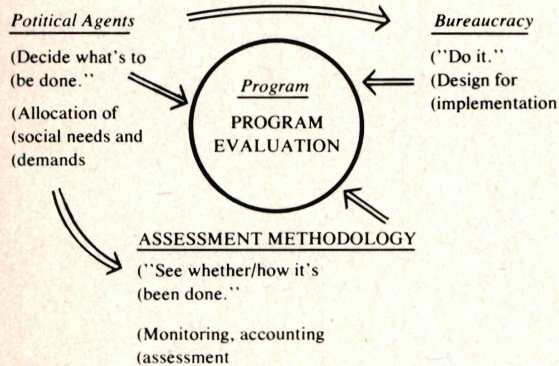
Effectiveness is not simply a technical concept handled in an abstract way by setting forth official program goals and measuring progress toward fulfillment of the stated goals. For any statement of effectiveness one must ask, "In whose terms?" It is necessary to analyze who might benefit and to recognize that there is not necessarily common interest among the different subgroups.

A list of who might benefit could be almost endless. It might include the following:

A. SPONSOR
  1. Strategic decision and funding
  2. Direction and coordination
  3. Monitoring and accounting

B. ORGANIZATION
  1. Management
  2. Program staff
  3. Other staff

C. PROGRAM EVALUATION TEAM
  1. Internal to organization
  2. External to organization

D. OTHER SERVICE DELIVERY ORGANIZATIONS
  1. With same program
  2. With different program
  3. With no program

E. CLIENT POPULATION
  1. Satisfied with program
  2. Not satisfied with program
  3. Quasi-clients; eligible but excluded

F. PUBLIC AT LARGE
  1. Welfare watchdogs and consumer groups; legal rights
  2. Cost-conscious community
  3. Therapy-treatment conscious community
  4. Social policy community: based on philosophic and academic ideas
  5. Social scientific community
  6. The politically interested
    a. Supporters
    b. Opponents
  7. Other groups

It is necessary to review potential benefits to such categories of partisans in terms of two types of uses:

A. CONSCIOUSLY INTENDED AND PUBLICLY ARTICULATED reasons for the study
  1. Agency accountability
  2. Basis of policy choice among alternatives
  3. Arbitration device for conflict resolution
  4. Rationalization of intrastructural growth
  5. Social surveillance and control of selected groups

B. POSSIBLY CONSCIOUSLY INTENDED, BUT NOT PUBLICLY ARTICULATED, reasons for the study
  1. Deflection of interest from sensitive areas
  2. Stalling tactic to buy time amidst embarrassment

3. Provision of data justification for elimination, cooling-out or hatchet decision
4. Justification for empire-building by means of official establishing of need, capability, etc.
5. New arena for academic games and gamesmanship
6. Partisan platform
   a. From client groups who might benefit
   b. To discredit policies of incumbents
   c. To demonstrate social inequality
   d. To develop a professional market
7. To attack or defend a previous evaluation

The point of this exercise is to keep the value issues prominent in the minds of the participants and to provide political information as a basis for understanding what is at stake in the evaluation, what coalitions can be formed or need to be neutralized, and what might be key points for intervention via the process of feedback and reporting, to improve the prospects for utilization of the findings from the evaluation.

I have suggested that the emphasis in the United States on managerial effectiveness may limit the utility of social program evaluation. In terms of program development, evaluation projects can be utilized to provide quick and reliable feedback in the sequential development of effective programs. Integrated evaluation thus becomes a mode of learning from experience rather than a final, concluding judgment.

Evaluation as an integral part of the project
Evaluation which becomes an integral part of the project or program improves (a) the *planning,* by emphasizing the need for operational objectives and relevant base-line data; (b) the *implementation,* by establishing functional feedback routines covering both costs and benefits and thus improving operational control of project activities; and (c) the *future development work* in related fields, by transmission of learning from experiences gained.

While one approach tends to neglect the possible learnings for and from other programs in favor of the continuing adjustment of and concentration on

the immediate program, the literature is gradually expanding attention to broader implications partly as it explores the issues of internal validity. In their most recent book my colleagues Epstein and Tripodi (3), who have consistently focussed on the utility of program evaluation for administrators and planners, provide carefully outlined designs for ''...formative evaluation strategies that are logical constructs for making inferences about program effectiveness and efficiency.'' (3, 114). By emphasizing the need for simplicity in comparisons and a concern for comparisons which have decision potential and involve program components which are manipulable (3, 149) they place the issues in a meaningful framework with both action potential and long-range implications. Such an approach represents a substantial contribution to increasing the utilization potential of social program evaluation.

But it would be inappropriate to my purpose to end this discussion on a technical note. As one author suggested in a recent paper, evaluation is a policy issue and its real benefit ''...may turn on the resolution of value conflicts.'' (13, 44) Tarter also suggests that evaluation performs a valuable function in illuminating the ''differences between the techniques of implementation and the values of an espoused policy.'' (13, 47) He urges that we ask, regarding any evaluation:

''What is the impact of the evaluation conclusions on the re-evaluation of policy, on the identification and examination of policy values, and on the creation of new alternatives for the application of those values in programmatic structures?'' (13, 52)

In a summary review article entitled ''The Failure of Social Programs'' James E. Prather and Frank K. Gibson (10) review evaluations of a number of different types of social programs, in education, criminal justice, mental health and organizational development. The report of overall negative findings is accompanied by the observation that these results have apparently had no substantial impact on policy makers.

''Faced with mounting political criticism, policy makers often establish program goals that are *politically* acceptable. Thus, educational decision makers em-

phasize student achievement primarily because it is the single criterion that most voters can understand and appreciate. If, in evaluating a particular program, student achievement scores do not respond favorably, the program is labeled a failure. But is it? . . . questions might have been asked that had little to do with achievement rates.

"Most input-output studies assign little weight to the political variable, yet it may have the strongest explanatory power of all variables considered.

"Finally, it is only fair to admit that the techniques available for evaluating program effects have progressed just slightly beyond the swaddling state. There exists a need for the type of policy analysis which has the range and perspective necessary for investigations and research into the *overall* (i.e., micro-level and macro-level) impact of programs and policy."

This summary citation has brought us full circle in our consideration of the relationship between program evaluation and the development of programs and policy. I have suggested that at present program evaluation plays a productive role in planning and program development, but almost no role in policy development, except via the consolidation of policy in program expression. I would further predict that the refinements will take place in program evaluation and program development but until we are able to forge a stronger link between micro-level and macro-level studies we cannot expect any substantial contribution from program evaluation to policy.

SELECTED BIBLIOGRAPHY

1. *Bennet, Carl A.,* and *Arthur A. Lumsdaine* (eds.): Evaluation and experiment. New York: Academic Press, 1975.
2. *Campbell, Donalt T.,* and *Julian C. Stanley:* Experimental and quasi-experimental designs for research. Chicago: Rand-McNally Course Publishing Co., 1965.
3. *Epstein, Irwin,* and *Tony Tripodi:* Research techniques for program planning, monitoring, and evaluation. New York: Columbia University Press, 1977.
4. *Etzioni, Amitai:* "Alternative Conceptions of Accountability: The Example of Health Administration." Public administration review, May/June 1975, pp. 279–286.
5. *Ferman, Lewis A:* "Some Perspectives on Evaluating Social Welfare Programs." The annals of the academy of political and social science (September 1969), 385, 143–156.
6. *Franklin, Jock L.,* and *Jean H. Thrasher:* An introduction to program evaluation. New York: John Wiley and Sons, 1976.
7. *Hage, Jerald,* and *Michael Aiken:* Social change in complex organizations. New York: Random House, 1970.
8. *Lind, Roger M.,* and *John E. Tropman:* Delinquency planning and community competence: A preliminary assessment. A research report to the Office of Juvenile Delinquency and Youth Development, U. S. Department of Health, Education and Welfare, Washington, D. C. March 1969.
9. *Marris, Peter,* and *Martin Rein:* Dilemmas of social reform. New York: Atherton Press, 1967.
10. *Patton, Michael Quinn:* Utilization-focussed evaluation. Beverly Hills/London: Sage Publications, 1978.
11. *Prather, James E.,* and *Frank K. Gibson:* "The Failure of Social Programs." Public administration review, September/October 1977, pp. 556–564.
12. *Rutman, Leonard* (ed.): Evaluation research methods: A basic guide. Beverly Hills/London: Sage Publications, 1978.
13. *Tarter, Rollin H.* Community mental health center program evaluation: Issues of policy and technique. Unpublished manuscript, The University of Louisville, 1978.
14. *Titmuss, Richard M:* The gift relationship. London: George Allen and Unwin, 1970.
15. —: "The Welfare State: Images and Realities." Social service review, Vol. xxxvii, No. 1, March 1963, pp. 1–11.
16. *Tropman, John E., Milan Dluhy, Roger Lind* and *Wayne Vasey:* Strategic perspectives on social policy. New York: Pergamon Press, 1976.
17. *Weiss, Carol H:* Evaluation research. Englewood Cliffs, N.J.: Prentice-Hall, 1972.
18. *Weiss, Robert S.,* and *Martin Rein:* "The Evaluation of Broad-Aim Programs: Difficulties in Experimental Design and an Alternative." The Annals of the academy of political and social science, September 1969, 385, 236–249.