



December 1974

## The Three Evaluations of Social Welfare Programs

Dean Harper  
*The University of Rochester*

Haroutun M. Babigian  
*The University of Rochester*

Follow this and additional works at: <https://scholarworks.wmich.edu/jssw>



Part of the [Social Welfare Commons](#), and the [Social Work Commons](#)

### Recommended Citation

Harper, Dean and Babigian, Haroutun M. (1974) "The Three Evaluations of Social Welfare Programs," *The Journal of Sociology & Social Welfare*: Vol. 2 : Iss. 2 , Article 13.

Available at: <https://scholarworks.wmich.edu/jssw/vol2/iss2/13>

This Article is brought to you for free and open access by the Social Work at ScholarWorks at WMU. For more information, please contact [maira.bundza@wmich.edu](mailto:maira.bundza@wmich.edu).



## THE THREE EVALUATIONS OF SOCIAL WELFARE PROGRAMS

Dean Harper  
Departments of Sociology and Psychiatry

Haroutun M. Babigian  
Departments of Psychiatry and Preventive Medicine

The University of Rochester  
Rochester, New York 14642

As new and innovative social welfare programs are being attempted, there has been an increased concern with evaluating the effectiveness of such programs. To what degree is a new program effective? For which kinds of clients is each type of program effective? What elements are crucial in a program which has been judged to be effective? These are just a few of the questions that evaluators would like to answer.

There is a large literature on evaluation research--some of it reporting or reviewing the results of specific evaluations (6, 8, 9, 10, 11) and some of it presenting general discussions, essays or models (1, 2, 3, 4, 5, 6, 7, 8). This literature outlines many important problems and suggests imaginative solutions to a number of those problems. However, in our judgment, some of the literature has viewed all evaluations as alike; insufficient attention has been given to the basic differences between different types of evaluation. The objective of this paper is to clarify this issue.

We argue that there are three distinct types of evaluation that can be articulated on the basis of the intent of the evaluation: (1) The evaluation is a test of some basic but general scientific hypothesis that underlies or motivates a proposed intervention. (2) The general scientific hypothesis is found or assumed to be valid, and the evaluation is an assessment of the effectiveness of a particular program that supposedly embodies the hypothesis. (3) After a particular program has been shown to be effective at some point in time, the evaluation is a continuing assessment of the effectiveness of the program. An awareness of these three types is important because the methods of evaluation are partly dictated by the intent of the evaluation.

These three types will be referred to, for the nonce, as evaluation-as-experiment (testing a scientific hypothesis), evaluation-as-assessment (judging the effectiveness of a particular program), and evaluation-as-monitoring (continuously examining the effectiveness of the program). Each of these will be discussed in turn.

## Evaluation-as-Experiment

As indicated above, each type of evaluation attempts to answer a distinct question. The first question is a question of "pure science."

Suppose that X represents some intervention which is thought to have some desired result, Y. Then, the evaluator asks: Is it true that X will, in fact, result in Y. Under what conditions Z will X result in Y? That is, he is interested in whether or not there is a direct and strong causal link between X and Y; and he is further interested in learning what other variables, Z, moderate or strengthen the causal relationship.

A clear example of this is found in medical research on new drugs. Suppose that a physician thinks that a particular new drug might reduce some of the symptoms of, say, schizophrenia. He needs to determine the effectiveness (Y) of giving the drug (X) to different types of patients ( $Z_1$ ), in different dosages ( $Z_2$ ), and in different settings ( $Z_3$ ). Obviously an experiment is the appropriate way to evaluate the drug.

Less clear examples of this are found in the evaluation of social programs. Consider a program directed to dealing with behavior problems (aggressive and rebellious behavior, withdrawal and the like) perceived in some school children. Suppose that a social worker concluded that the problems of these children were manifestations of family problems. Since the social worker could not work with individual families he decided to conduct small group therapy sessions, with, say five or six families (i.e., pairs of parents) in each group; his decision was based on the conception that through discussion with others, each set of parents might gain some insight into their own conflicts and problems and from that insight alter their interaction with and treatment of their children. Thus, the general behavioral proposition that would motivate this approach is that discussions with others who have similar problems leads to insight into one's own problems which, in turn, leads to change in one's interactions with others which, in turn, leads to an alleviation of the behavioral problems in those others.

It may be that group therapy with parents has, in general, little impact on the behavioral problems of offspring. It has some effect but much less than, say, group therapy with the adolescents. Or, alternatively, the behavioral problems of adolescents may co-occur with failures in school--academic and social failures--and further, the behavioral problems are a consequence of the failures. That is, educational activities which permit these adolescents some experiences of success have the consequence of altering their other behavior which was defined as problematic or psychopathologic.

Here, then, are three distinct principles (group therapy with parents, group therapy with adolescents, experiences of success) on which remedial programs could be based; there are many others. These principles or propositions have the status of scientific hypotheses; as such they need to be tested in a rigorous experimental fashion.

An experimental investigation could, of course, be devised whereby families are randomly assigned to one of several treatment groups (no therapy or treatment vs. group therapy with parents vs. group therapy with adolescents vs. experiences of success vs. group therapy with parents and with adolescents, and so on), with the outcome measure being some set of systematic observations on the behavior of the adolescents. And these should be tested in a variety of contexts: (1) varying economic and educational background of parents, (2) type of community, e.g., racially mixed or not, working class or middle class and so on, (3) type of behavioral problems of children.

The point is not to show whether these treatments are effective or not but to obtain some measure of relative effects. Thus, it might be the case that, in general, group therapy has some effect but relative to experiences of success it is of little importance. That is, if children have some rewarding academic or social achievements in school, then this has a greater impact than gaining insight into their own problems.

There are few instances of planned social intervention that have been preceded by careful experiments conducted in a variety of situations. The problems of the day are so pressing that welfare personnel clearly cannot wait until definitive experimental results are obtained; they must create plausible programs and proceed in a trial-and-error fashion. However, at the same time, some experimental investigation should be conducted. The experimental investigation constitutes one type of program evaluation.

#### Evaluation-as-Assessment

The question asked in the first kind of evaluation is a scientific question; this does not imply, however, that the question underlying the second kind of evaluation is unscientific. Rather, the second question is a question in applied science.

The second question assumes that the first, the question in basic science, has been answered affirmatively. That is, it has been reasonably demonstrated that under conditions Z a specified X has, in general, a desired result Y.

Then the questions are the following. For a particular instance of the intervention, i.e., a program of planned change that is operating in some particular community: (1) do the conditions Z exist in the community? (2) does the program embody the specified X, the agents designed to bring the change? (3) are the desired consequences, Y, achieved?

An example from medicine may make this clear. A surgeon may have discovered and reported that for patients of certain ages and constitutions, a particular surgical technique is effective in correcting certain heart defects. Over a period of time this surgical technique gains wide use. Another surgeon might then be interested in evaluating the effectiveness of this surgical procedure in treating patients with these heart defects. The conditions, Z, then, are the characteristics of patients for whom the technique is useful. The change agent, X, is the surgical technique. The desired objective, Y, is the correction of heart defects.

In conducting the evaluation the surgeon might learn that Y is not achieved; a significant number of heart defects are not being corrected. Several possibilities may account for this. The conditions Z do not exist; the surgical technique is being used on inappropriate types of patients. Alternatively, where the operation is attempted, it is not being performed in the specified way; that is, physicians throughout the various hospitals are not performing it in the way it was originally contrived. In our formalism, the specified X is not embodied in the medical care programs. Still a third possibility is that the original research leading to this surgical technique was in error.

There is considerable difference between medical care programs and social welfare programs; one difference, of course, is that social science has not supplied general research findings that clearly dictate one program rather than another. As we have indicated above, social scientists should direct some of their efforts to obtaining this knowledge.

If the necessary basic research had been done, which had demonstrated that some principle is effective in ameliorating some social problem, then programs using that principle could be assessed. For example, suppose that the definitive experiment had been conducted which demonstrated that group therapy with parents and school success of adolescents was significantly more effective than either alone and that group therapy with adolescents did not have much additional impact. And, suppose that this definitive experiment had articulated rather precisely the nature of the group therapy and the experiences of success so that the program could be copied easily in appropriate settings. Finally, suppose that the program was begun in Community A. Now an investigator in Community A wants to evaluate the effectiveness of this program as it is conducted in his particular community. Since the principles underlying the program have been experimentally validated, he is not going to replicate the scientific experiment. Rather he wants to know (i) are the principles and concepts of this program satisfactorily embodied in the particular activities of Community A? and (ii) is there a reduction in the behavioral problems of adolescents?

This investigator may learn that the desired results are not achieved; there is no reduction in the frequency or degree of behavioral problems of adolescents. But because he asked and attempted to answer the first question above, he can suggest a cause for its ineffectiveness; he learns that the program did not involve experiences of success. Academic tasks that the adolescents could excel in were not devised; there were only a few abortive social affairs, none of which resulted in social successes for these young people. Hence, the investigator is able to indicate why the program failed.

Although the first kind of evaluation requires an experiment, this is not necessary in an assessment; this does not mean, however, that it is easier. In the experiment, some outcome (e.g., reduced behavior problems) in experimental groups is compared with that outcome in the

control groups. In the assessment, however, some previously established standard corresponds to the control group; that is, the investigator compares the results obtained or observations made in the evaluation with the specified standard. The conclusion about the effectiveness of a particular program in a particular setting depends on whether or not its results equal or exceed the standard.

Perhaps we have overdrawn the distinction between evaluation-as-experiment and evaluation-as-assessment, at least as far as evaluation research is practiced, but we do believe that this distinction reflects an important problem that has not been sufficiently appreciated by some of those conducting evaluations.

### Evaluation-as-Monitoring

The second kind of evaluation can merge into or become a third kind: monitoring. Less needs to be said about this third kind; however, it is necessary to keep the second and third distinct--both conceptually and in practice.

Like an assessment and unlike an experiment, monitoring is directed to a particular program--rather than to the class of all such programs. The assessment can be viewed as a discrete investigation with a beginning and an ending. On the other hand, the monitoring of a program is a continuous evaluation. It can be likened to the quality control techniques of the industrial engineer.

The assessment of a program rests on the experimental evidence that dictates the principles underlying the program. Thus, logically, assessment follows experiment. Likewise monitoring follows assessment; the monitoring of a program rests on an investigation that revealed that the program was effective during a previous period of time.

Suppose for example, the administrator of a school mental health program in Community A has had his program assessed with the result that it is functioning fairly effectively. The next step he may wish to take is to devise some means of continuously or periodically evaluating its functioning; that is, he begins some procedure of periodically making observations of the program in operation. These observations may lead to immediate changes in its operation. Monitoring--the systematic and periodic observation--is a "servo-mechanism" in the program; it feeds back data which leads to self correction.

### Conclusion

The evaluation of the effectiveness of social welfare programs is an important and necessary activity. Among the many considerations--political, social and scientific--we have focused on one: the methodological issues of types of evaluations. We have argued that there are three types of evaluation--all of which are necessary. In making an

evaluation, an investigator should know what he is and is not doing; he should see the formal relationship between his own research activities and that of others, which though seemingly similar to his own, may be different in some significant way. If an evaluator can locate his own attempt at evaluation in terms of the structure of evaluation research that has been outlined in this paper, then his attempt will be both more efficient and effective.

#### References

1. Borgatta, E.F. "Research Problems in Evaluation of Health Service Demonstrations." The Milbank Memorial Fund Quarterly, 44 (October 1966): 182-200.
2. Brennan, John G. "Evaluation of Community Mental Health Programs: A Pragmatic and Theoretical Conceptualization." Social Psychiatry, 6 (September 1971): 116-121.
3. Collins, J.A. Evaluative Research in Community Psychiatry, Hospital and Community Psychiatry, 19 (April 1968): 97-102.
4. Donabedian, A. Evaluating the Quality of Medical Care, Milbank Memorial Fund Quarterly, 44 (July 1966): 166-203.
5. Fischer, Joel. "A Framework for the Analysis and Comparison of Clinical Theories of Induced Change." The Social Service Review, 45 (December 1971): 440-454.
6. Herz, Marvin I.; Endicott, Jean; Spitzer, Robert L.; and Mesnikoff, Al. "Day versus Inpatient Hospitalization: A Controlled Study." American Journal of Psychiatry, 127 (April 1971): 1371-1381.
7. Jaeger, B.J. "Evaluating National Hospital Policy: A Systems Approach." Health Services Research, 7 (Spring 1972): 11-22.
8. Schulberg, H., Sheldon, A., and Baker, F. Program Evaluation in the Health Fields. New York: Behavioral Publications, 1969.
9. Shireman, Charles H.; Mann, Katherine Baird; Larsen, Charles; Young, Thomas. "Findings from Experiments in Treatment in the Correctional Institution." The Social Service Review, 46 (March 1972): 38-59.
10. Spergel, Irving A. "Community-based Delinquency-Prevention Programs: An Overview." The Social Service Review, 47 (March 1973): 16-21.
11. Wilkinson, Kenneth P.; and Ross, Peggy J. "Evaluation of the Mississippi AFDC Experiment." The Social Service Review, 46 (September 1973): 363-377.