Value Assumptions Underlying Evaluation Research

Thomas J. Cook
Frank P. Sciolli Jr.

Follow this and additional works at: https://digitalcommons.coastal.edu/jops
Part of the Political Science Commons

Recommended Citation
Available at: https://digitalcommons.coastal.edu/jops/vol4/iss2/4

This Article is brought to you for free and open access by the Politics at CCU Digital Commons. It has been accepted for inclusion in Journal of Political Science by an authorized editor of CCU Digital Commons. For more information, please contact commons@coastal.edu.
Value Assumptions Underlying Evaluation Research

THOMAS J. COOK  
*University of Illinois at Chicago Circle*  

and  

FRANK P. SCIOLI, JR.  
*National Science Foundation*

Introduction

An important area in the study of public policy has been research directed at assessing the consequences of policy action. Generally conducted under the heading of program evaluation, the research attention has centered both on the development of appropriate methodologies for evaluation analysis as well as the measurement of the results of substantive policy actions. Relatively little attention, however, has been directed at the various assumptions which underlie much of the research in the area of program evaluation. In this paper we will address several of the major assumptions and discuss their relevance to the conduct of an evaluation analysis. In particular, we will discuss the potential impact of these assumptions on issues such as the scope of the evaluation, the choice of evaluative criteria, and the type of research design employed. While we do not assume that the list of assumptions discussed is exhaustive, we do intend that it be provocative and, we hope, representative of the assumptions underlying program evaluation in various substantive contexts.

Terminology

In general terms, program evaluation may be conceptualized as the research enterprise concerned with analyzing the relationship between program intentions and program performance. More specifically, it measures the extent to which a program action was successful in achieving its intended goals or objectives.\(^1\) In principle the output from a program evaluation serves as input (i.e., feedback on program results) for decision-making about future resource allocations (e.g., to continue, modify, or discontinue a program). One can easily find numerous candidates for a program evaluation. Will a Volunteer-in-Probation program significantly decrease the probability of recidivism among adult misdemeanants? Will a high-intensity street lighting program significantly decrease the incidence of on-street assaults in a given

---

area? Will an income maintenance program be an incentive for finding employment? Will an “operation identification” program significantly increase the recovery of stolen property by those victimized? These are all potential targets for a program evaluation. Although the programs differ, the general research problem is the same: determining the specific effects which may reasonably be attributed to a given program activity.

Important to an appreciation of this type of research is an understanding of the assumptive behavior which guides much of the evaluative research. By assumptions, we mean those aspects of evaluation research which are “taken for granted,” or at least not tested directly, and yet may have important consequences for either the social utility of the program or the validity of the evaluation conclusions. For example, one might adopt a narrow view of program impact, and, consequently, conduct observations on a limited set of impact indicators. This narrow view may reflect the assumption that only positive, or beneficial, consequences will result from a particular activity. Much of the Volunteers-in-Probation movement is characterized by this viewpoint, that is, if a person volunteers his time, he or she must be doing some good. In many cases, there is a reluctance to admit the possibility of negative effects resulting from volunteer-probationer interaction, such as in one-to-one counseling. Thus, the assumption of “only positive results” may limit the types of indicators selected for measuring program impact and set boundaries on the scope of program consequences which could be measured. In this paper we will focus on various assumptions dealing with the values, or subjective preferences, which may influence the conduct of a program evaluation.

Value assumptions got to the very heart of the evaluation enterprise and pertain to the basic premises upon which the research is conducted. They raise issues about the principles which should guide evaluation research such as the “legitimate” purposes of evaluation research and the “proper role” of the program evaluator. In other words, what questions ought to be asked and how should they be investigated? Value assumptions are important, as they often establish the boundaries for the evaluative effort in that they set limits as to the scope and nature of the questions asked vis-a-vis the determination of program “success.” In most cases they are implicit, rather than explicit, and the prescriptive guidelines suggested by the assumptions are followed in the absence of any justifying evidence, empirical or otherwise. As the discussion below will point out, there is also somewhat less than a perfect consensus on the “best” set of prescriptive guidelines to be followed in a program evaluation.
Value Assumptions

The first assumption deals with program evaluation as a competitor for resources in the over-all budgetary allocation of a program. The pro-evaluation assumption is that every program can, and should, be evaluated. Moreover, it is assumed that a program, even before it is funded, should have built into it the mechanism for determining whether or not the program was a success. An example of this is the provision in Law Enforcement Assistance Administration (LEAA) funding which requires that funded programs must have an evaluation component. The trade-off here is that money spent for evaluation is not available for providing program service, or treatment. Thus, the opportunity-cost of denied treatment is balanced against the desire for information about the extent to which a program is achieving its objectives. It is assumed in this type of budgetary situation that the value of the program evaluation (expressed in terms of providing more efficient and effective service in the future) outweighs the short-run deprivation of individuals who would otherwise receive the service contained within the program.

A corollary point of a methodological nature deals with the use of control groups in program evaluation. The argument here is that a more valid evaluation can only be obtained when a treatment group is compared against a group which did not receive a particular treatment, in order to have a comparative baseline from which to assess the impact of a particular program. We label this the “control group dilemma” in that a control group is denied a service that “may” be of great benefit in order that the researchers maximize the value of obtaining a “valid” evaluation.

The counter argument contends that it is wrong to deny a potentially beneficial service to a group of people merely to achieve a desired level of methodological sophistication. One often hears this argument from personnel responsible for administering a program. Their contention is that “we’re too busy helping the people in the program to take time out for evaluation, and besides it’s wrong to deny people a service they really need.” The rebuttal, of course, is “how do you know the program is effective, and meets the needs of the people, unless you’ve systematically evaluated it?” Thus, we have a conflict in values — allocation of resources to provide service vs. the allocation of resources to determine service effectiveness.

A related dispute is relevant to the contention that every program can, and should, be evaluated. One position is that some evaluation is better than none at all, regardless of the data limitations present in the evaluation setting. The converse position is that of foregoing an evaluation of a program where the design of the evaluation is severely constrained by the available data and there is high probability of inconclusive, or misleading, results. The debate of
Campbell and Erlebacher vs. the Westinghouse Group is illustrative of this point. Campbell and Erlebacher contend that the Head Start evaluations had such severe methodological problems, especially with regard to the regression artifact threat, that the evaluations should not have been undertaken since the conclusions were of such dubious validity and, hence, potentially misleading. The rejoinder to this argument was that while the regression artifact problem was present in the data, the program was of such importance that an evaluation of it, even if only partially valid, was required for future policy decisions. The conflict here is between the preference for only highly valid program evaluations vs. the acceptability of “satisficing” evaluations under conditions of limited control over the research setting. The former position argues that a “bad” (i.e., methodologically suspect) evaluation is worse than no evaluation at all, whereas the latter position contends that the validity of an evaluation is a matter of degree and that some information about program success is better than no information at all.

A third major area of value conflict lies in the emphasis upon the “Scientific Method” as the correct model for program evaluation. The dispute often centers upon the quantification of program objectives and measures of program success. The conflict becomes readily apparent when one contrasts the views of the program personnel with those of the academic researcher. An example of this can be found in the area of volunteer programs in courts and corrections. The academic researcher will call for objective measures, such as the probability of recidivism, whereas the volunteer will stress more subjective impressions like “the person feels better about himself and is much more able to relate to others.” In our research in this area, we have run across statements such as “It’s impossible to quantify the effectiveness of one-to-one counseling because it is such a personal matter.” A further contention is that quantification tends to force concepts, such as program objectives and effectiveness criteria into unrealistic and impractical categories which lose much of the “essence” of the program as it operates and affects people’s lives. The converse position is that these types of qualitative impressions are not objective indicators and are totally limited to personal perceptions and subjective interpretations, hence, they do not provide unambiguous measures of program goals or the extent to which these goals have been achieved.

The conflict seems to center on the preference for "analytical objectivity" vs. the preference for a more personalized approach which evaluates a program in terms of the qualitative differences resulting from program participation.

A variant of this conflict is very pronounced in the area of "human experimentation." One argument is that you can't experiment with people's emotions, feelings, beliefs, etc., but rather these must be approached within a more subjective inter-personal mode of analysis. The converse position is that experimentation is the most valid means for testing hypothesized causal relationships between program activities and measurable changes in program-relevant behavior. One might contend this conflict goes even deeper and relates to the point made above concerning the "control group dilemma" whereby services are withheld from a group in order to provide a baseline from which to measure program effect. The anti-experimentalist would argue that it is wrong to withhold a needed service to satisfy a methodological requirement, such as random assignment to treatment and control groups. Besides, the impersonal nature of the random assignment procedure negates the essential "humanity" of the program. The pro-experimentation person would counter that, in the more realistic situation of scarce resources and competing claims for resources, the random assignment procedure is actually the "fairest" means of allocating the available resources since everyone has an equal chance of being selected to receive the program service.

A fourth area of value conflict concerns the "proper" role of the program evaluator. This assumption posits that the primary job of the program evaluator is to measure the extent to which a program achieves its objectives, and not to question either the social utility of the objectives, or the values upon which they are based. This view conceives of the program evaluator as a "neutral" analyst whose job is to evaluate program success in terms of the objectives as stated by the responsible program agency. The question here is not only can the evaluator be a completely neutral observer but also should he be a neutral observer? An example might point up the potential dilemma here. It could be argued that the most efficient way to reduce the probability of a person convicted of murder being a repeater of that crime is to implement a program of capital punishment. Likewise, one might argue that an effective way of dealing with convicted rapists would be to castrate them, à la the Billy Graham suggestion of a few years ago. Is the proper role of the program evaluator merely to assess the extent to which these types of programs achieve their objectives or does he have a responsibility to question the values which

---

underlie the programs? That is, should he choose between the value of a short-run reduction in criminal activity vs. the value of preserving human life and related capacities in the hopes of future rehabilitation?

To consider another example: Suppose a researcher has been commissioned to evaluate a program which has the primary objective of increasing the absolute reading level of black children located in several ghetto kindergarten classes. Suppose, however, that his own preference is for a program designed to reduce the gap between the reading level for black children and that of middle-class, white children located in selected suburban schools. In the latter instance, the performance levels of the white children would constitute the standard by which the program is evaluated rather than the pre-program levels of the black children. The situation confronting the evaluator is fairly clear. Does the evaluator accept the objective as set forth in the official program statement and use this as the standard for evaluating the program, or does he exercise an independent judgment in challenging the ultimate social value of the objective and become an advocate for a different set of program objectives and, consequently, a different set of evaluative criteria? Moreover, should the evaluator employ his preferred set of evaluative criteria even if it may reduce the acceptability of his recommendations by the sponsoring agency and hinder their chances for future policy implementation? As can be seen, the choices made by the evaluator in these types of situations may have profound consequences for the conduct of the evaluation, the kind of information it produces and the possibility of affecting future policy.

A corollary point can be made regarding the perspective from which the program is evaluated. In employing the agency’s program objectives as the performance standard, the researcher is assuming that these objectives are consistent with the needs of the clientele supposedly served by the program. The extent of consistency, or inconsistency, between official program objectives and client needs is, of course, very important if one evaluates the program from the perspective of the client. Should the program evaluator merely assume that a high degree of consistency exists, or should he seek evidence concerning this relationship? If he finds a high degree of inconsistency, what is his proper course of action regarding the conduct of the evaluation? In other words, does he become a spokesman for the clients of the program in developing effectiveness criteria which reflect their extent of need satisfaction, or does he become an agency spokesman by evaluating the program in terms of the objectives that the agency has identified and upon which the agency wishes the program to be evaluated?

The point on the proper role of the evaluator introduces another area of disagreement which has received considerable attention in the literature.
This involves the issue of whether or not the programs should be evaluated by people inside or outside of the agency administering the program. One position argues that the program should be evaluated by people who are intimately involved in the administration and implementation of the program because they have the greatest amount of familiarity with the program, understand the idiosyncratic features of program implementation, and, consequently, are in a better position to evaluate the success of the program. The converse position is that people involved in administering the program may have a vested interest in producing evaluations which show that the program is a success. Thus, their involvement in, and commitment to, the program may bias their judgment and preclude their being able to render an "objective" evaluation of program achievement. This position argues that the evaluation should be performed by a person from outside the administering agency who has the necessary "independence" to conduct an impartial evaluation. The outside evaluator would, in effect, lend an aura of legitimacy to the evaluation by his non-involvement (at least formally) in the administration of the program. There is no easy solution to this dilemma and numerous people fall on both sides of the question. The choice rests upon the assumptions made concerning the relative value of an "impartial" outside evaluation vs. an inside evaluation by people actively involved in the program who have a "fuller understanding" of the program's operation.

The final consideration involves an area which, perhaps in large part, affects all of the previously discussed issues. This entails the criteria for selecting programs to be evaluated. Since, obviously, not all social programs have been, or will be, evaluated, choices are made to evaluate some programs and let other programs go unexamined. In the final analysis, these kinds of choices most directly relate to determining the value of program evaluation as a generator of decision-relevant information, since the general utility of evaluation will be largely determined by the type of programs evaluated. Several questions may be raised about what types of choices are made and the values they reflect. Is a program selected for evaluation because it does not involve a politically sensitive area and it would therefore be relatively easy to obtain the necessary cooperation for the evaluation? Is a program evaluated because it involves an area which has readily available data? Is a program selected because the researcher is confident that his recommendations would actually affect future policy? Is a program not evaluated because it has been in


existence for a long period of time and has developed a powerful supportive constituency? These, and other examples which could be mentioned, all represent possible criteria which might go into the decision to select a program for evaluation. As such, they reflect both the preferences of the researcher and/or his perceptions of the realities endemic to the potential evaluation setting. In any case, the selection criteria not only serve to determine which programs will be evaluated, but may provide an indication of the future prospects of a program even before the evaluation results are obtained in that they may reflect the predominant values, or preferences, at a given point in time.

Conclusion

In this paper we have had the luxury of raising, and not being forced to resolve, a number of important issues involving conflicting value assumptions relative to evaluative research. Perhaps, as some would argue, these issues escape a final resolution, and are best left to the philosophers among us. We disagree. We feel that they are a legitimate concern of the policy analyst and must be addressed both in terms of explicating the assumptions underlying an evaluation and assessing their impact upon policy recommendations. We trust that this paper represents a positive step in that direction.