



Social Forces, University of North Carolina Press

The Multi-Goal, Theory-Driven Approach to Evaluation: A Model Linking Basic and Applied Social Science

Author(s): Huey-Tsyh Chen and Peter H. Rossi

Source: *Social Forces*, Vol. 59, No. 1 (Sep., 1980), pp. 106-122

Published by: [University of North Carolina Press](#)

Stable URL: <http://www.jstor.org/stable/2577835>

Accessed: 28/05/2011 10:04

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=uncpress>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Social Forces, University of North Carolina Press and University of North Carolina Press are collaborating with JSTOR to digitize, preserve and extend access to *Social Forces*.

<http://www.jstor.org>

The Multi-Goal, Theory-Driven Approach to Evaluation: A Model Linking Basic and Applied Social Science*

HUEY-TSYH CHEN, *University of Massachusetts*

PETER H. ROSSI, *University of Massachusetts*

ABSTRACT

A source of one of the more serious problems of evaluation research is the typical finding that evaluated programs have little or no effectiveness. Recently, claims have been made that the problem of no effect may be attributed to the inadequacy of current evaluation methodology. In this paper, the reasons for no effect in current program evaluation are analyzed theoretically. In order to cope with this problem, an alternative evaluation method, the "multi-goal, theory-driven" approach, is proposed. The principles, procedures, and justification of this new evaluation approach are discussed. Finally, the advantages of applying the multi-goal, theory-driven approach to program evaluation are illustrated.

"No Effect" Outcomes: Findings or Method?

In evaluation research it appears that nothing succeeds like failure. While more and more agencies and policy makers turn to evaluations for information on social programs, evaluation researchers find increasingly less and less about which to be sanguine in our current social programs. In one field after another, evaluation researchers find that the programs in place or contemplated have few or no effects of the sort intended by their designers. We now know that rehabilitation efforts fail to reform prisoners; that more money and curriculum changes fail to increase the teaching abilities of schools; that poverty is scarcely to be alleviated by counselling the poor; that housing, when improved, leads to no corresponding changes in the quality of lives of residents; and manpower training programs scarcely improve the

*The preparation of this paper was supported by NIH grant #5T21MH13740-04 of which Rossi was Principal Investigator.

© 1980 The University of North Carolina Press. 0037-7732/80/010106-2\$01.70

employment chances of graduates; and so on, through a long litany of programs that when tested against fair measures of success were found to be ineffective or failures.

From this state of affairs, two divergent conclusions have been drawn: on the one hand, the findings of “no effects” have been interpreted as reflecting the failures of the social engineering efforts embodied in the programs (Rossi), an assertion that we simply do not know enough to be able to design programs that are effective in achieving given goals or that can be effectively implemented on a mass scale. On the other hand, some commentators (Scriven; Weiss and Rein) have insisted that the failures of programs reflect the failures of evaluation methods, insisting that the research paradigms employed contain built-in obstacles to finding effectiveness.

There is some substance to each of these interpretations. Social programs of a fairly wide variety have been tried with various levels of effort. It is hard to believe that most, let alone all, such programs have been badly designed, poorly planned, or implemented with such indifference that almost all are worthless. This suggests that we have been less than effective in detecting the impacts of such programs. At the same time, it is also hard to believe that the research paradigms employed in evaluation have been defective. Evaluations have become increasingly sophisticated as the technical problems of evaluation research have drawn the attention of some of the more creative methodologists from all of the social sciences. Research designs appear to be at least adequate to the task.

Furthermore, calls to relax design rigor are also suggestions that we ought to admit more Type I errors—false positives—a strategy that implicitly claims it is better to waste social resources on ineffective programs than to have no programs at all. As Gilbert et al. have shown, the more rigorous the research design, the more likely the evaluation research is to find no effects.

It is our thesis that both interpretations described above are off the mark. In our estimation the problem lies in the articulation of research design and program design. Evaluation researchers have not adequately mapped social programs on to the research designs that are used. In this interpretation, there is nothing wrong with the formal structure of conventional research paradigms, nor are there necessarily serious defects in the programs. Rather the problem lies in the extent to which programs have been properly interpreted in the designing of evaluation researches. We will also suggest an approach to the articulation of social programs and evaluation research designs that we believe will improve the sensitivity and responsiveness to policy needs.

There are two critical elements in the approach we suggest. First we contend that *every program has some effects*, possibly only trivial, an assertion based on the idea that any purposive social action of any appreciable mag-

nitude does disturb the social system to which it is directed. Second, we assert that *a priori knowledge and social science theory can adequately anticipate the effects that a given social program can be expected to have*. These anticipated knowledge or theory-based effects can be assessed in the course of a properly designed evaluation effort.

Of course, we are not alone in questioning the adequacy of evaluation research, as conventionally defined. Shirley Angrist has noted the limitations of current evaluation research in detecting side effects. Irwin Deutscher has coined the phrase "goal trap" to designate the pitfall into which evaluators may fall when exclusive attention is paid to official program goals. Even more specific has been Michael Scriven's suggestion that evaluation research should proceed in a "goal-free" fashion. Scriven's approach is sufficiently intriguing to be worth considering in detail, especially since his suggestion is in some respects similar to our own.

Scriven's proposed goal-free evaluation approach also was apparently born out of the frustrations of finding no effects in one field experience after another. But he also noted that, although program-defined, intended effects usually failed to appear, unintended effects, unanticipated by program designers, often surfaced. Programs were having some effects, but not those that related to the usually ill-defined and vague goals set by program designers and administrators.

As a remedy, Scriven proposed that evaluators ignore—indeed, even avoid knowledge of—program design and administrator-defined goals and concentrate on those effects that appear to be occurring. Scriven is unfortunately somewhat vague about how the evaluator is to decide which among the possibly infinite number of potential effects are to be studied by evaluators. Events, processes and outcomes are not obvious; they can only be discerned with appropriate conceptual schemes that provide guides for observation. Undisciplined ferretting about for differences between treatment and control groups, for example, might maximize Type I errors, confusing chance generated differences for program related ones.

The approach proposed in this paper resembles Scriven's goal-free evaluation ideas in emphasizing that programs have some effects that do not necessarily coincide with the intentions of designers and administrators, but differs from that approach by suggesting ways in which such effects may be discerned. Our approach entails *defining a set of outcomes as potential effects of a program, some given by the official goals of the program and others derived from social science knowledge and theory concerning the subject matter in question*. This multi-goal, theory-driven approach is described in the remainder of this paper.

THE MULTI-GOAL, THEORY-DRIVEN EVALUATION APPROACH

With few exceptions,¹ most evaluations follow this pattern. Program designers and administrators construct an intervention that is intended to fulfill certain stated goals. For various reasons, statements of program goals tend to be vague and/or stated in terms that do not lend themselves easily to precise measurement. Such expressed program goals usually are treated by evaluators as indicating where to search for program effects. Since measurability and specificity are greatly emphasized by evaluators, in practice one is able to define only a few measurable, specific outcome variables (goals). The set of such outcome variables is typically quite narrow compared to the original social problem that the social program aims to alleviate. For example, the evaluated goals of Performance Contracting Experiment when specified turned out to be increasing disadvantaged students' reading and math scores, the evaluated goals of Head Start turned out to be increasing cognitive and affective development, etc., definitions which are considerably narrower than the stated goals articulated by program designers.

An instructive example is an analysis of the Community Mental Health Center's program of National Institute of Mental Health, in which Wholey and his associates suggested that all of 46 branch goals and 5 out of 6 program goals were not evaluable "because they were not stated in measurable terms" (92). The sole surviving measurable program goal of CMHC program according to Wholey is "economically viable CMHC's, independent of federal support" (92), certainly much narrower a focus than the program's intention.

This current evaluation approach may be labelled the "official-goal-fixed approach," since the evaluators mainly focus their attention on a few narrowly defined measurable effect variables picked from among official proposed program goals, and assess whether the program treatments affect those narrowly defined outcome variables. If a treatment achieves those effects, then a program is declared a success; otherwise it is a failure.

The official-goal-fixed evaluation paradigm often (if not usually) finds that a given program does not work. That is, little evidence can be found that those measurable goals, filtered out of the full set of goals stated by program personnel, are achieved by the treatment involved. Clearly the problems with the conventional paradigm are not to be remedied by using rigorous research designs since there would be even less chance of finding effects using more powerful designs. Rather the problem may be whether the official-goal-fixed paradigm is appropriate.

The bases for this argument are as follows:

1. *Social Reforms Are Not Social Science Based*

A crucial weakness in current social reforms is that they are ordinarily

divorced from social science understanding of social problems. In part, this occurs because policy makers and program designers do not look to social scientists for help in the design phase, and in part because, with the possible exception of economics, the social sciences have yet to develop an adequate set of theories that are relevant to social problems. In addition, it is not often recognized that at least two kinds of social science theory are necessary: theories that model the social problem in question and theories that model programs. Although the two sets of theories must be consistent one with the other, it should be recognized that treatments are not directly derivable from understanding of problems and vice versa.

Both the diagnosis of social problems and the design of treatments are now the provinces of policy makers and program designers, both often trained in law schools and with little social science training. As a consequence, programs are often designed based on conventional, commonsense understandings both of social problems and of treatments. Conventional evaluation paradigms that accept program goal and treatment conceptions that derive from policy maker and administrator definitions are thus hitched to conventional, commonsense notions that may be quite wide of the mark. The results of this practice are that, on the one hand, programs evaluated by the conventional paradigm often show no effects. That is not surprising since whether the treatment works as intended is not well known in the first place from previous research or viewed from a knowledge or theory base. Yet there may be some effects other than official intended ones that might have been uncovered if the evaluation had looked for them.

These arguments seem to suggest that it is fruitful for evaluators to identify a range of outcome variables including not only the administrators' intended program goals, but also those that might be indicated by the evaluators' understanding as to outcomes most likely to be influenced by the treatment. Existing social science knowledge may not be enough to tell us which specific outcomes will be precisely affected by the program treatments, but it is likely to be enough to tell which outcomes are *most likely* to be influenced by the treatments of the program.

2. *The Misleading Range of Administrators' Program Goals*

The program goals prescribed by policy makers and administrators are not necessarily the effects which are most likely to be achieved by the treatments delivered by a program. Administrators' program goals appear to be selected under two kinds of criteria: the first is desirability, the second is possibility. It often happens that the administrators pick goals more on the basis of desirability or hope than possibility or understanding. Accordingly, official-goal-fixed approach evaluators who use administrators' statements of program goals as the limits within which to search for measurable effect variables may be on the wrong track in the first place. The range of admin-

istrators' program goals may lead official-goal-fixed approach evaluators to use as measurable goals ones which are not likely to be affected by program treatment. If this is the case, the findings of little or no effects through using official-goal-fixed approaches are understandable.

Although the above reasoning helps to explain why so often no effects are detected, it does not provide a justification for searching for effects that were not intended but which were nevertheless consequences of the treatments administered in the program. That rationale is provided by the depths of our desperate need to know more about social problems and treatments. A conventional paradigm evaluation that provides us with information on whether a treatment works in some narrow range of intended measurable effects provides only a minimum amount of potential knowledge. What the treatment in fact does—what effects occur, intended *and* unintended—can enlarge our knowledge base about both the social problem in question and adequate treatments. This additional knowledge can be the basis for the construction of better programs and better social science understanding of programs and social problems.

An illustration may be appropriate at this point. In a randomized controlled experiment testing the efficacy of extending unemployment insurance coverage to prisoners released from state prisons in Georgia and Texas, a straight ANOVA analysis indicated that the treatments had no effect on re-arrest rates in the year following release (Rossi et al.). A theoretical analysis of the treatment indicated the possibility of two counter-balancing effects, one in which the unemployment benefits reduced arrests by providing income during the transition between prison and full integration into civil life and another which raised arrests because the unemployment benefits provided a work disincentive and increased unemployment and hence arrests among those ex-felons who were in the treatment groups. Social science theory derived from both microeconomics and sociology provided the analysts with the means to go beyond the no-effects finding to specify a model of the experiment that more adequately accounted for outcomes. The theory-based model also made it possible to propose better treatments that would accomplish the aim of reducing re-arrest rates for released prisoners without producing the undesirable side effect of increasing unemployment.

Under the multi-goal, theory-driven approach, the evaluator's role is different than in the conventional official-goal-fixed approach. In general, conventional evaluators passively accept the goals they are asked to evaluate, with social science knowledge and theory playing a minor part in the evaluation process. In contrast, in our suggested approach, the evaluator should actively search for and construct a theoretically justified model of the social problem in order to understand and capture what a program really can do for a social problem—social science knowledge and theory become crucial in the evaluation process.

Since program evaluation is usually pursued under contracts, it requires the consent of program administrators and/or policy makers to deal with phenomena which lie outside officially intended goals. One might hope that it would not be difficult to obtain permission to undertake multi-goal, theory-driven evaluation, since this evaluation provides program people a much more profound understanding of the potentiality of a program than the traditional approach.

One of the major features of the approach suggested here that should prove attractive to program administrators and policy makers is that the multi-goal, theory-driven approach promises to provide a much more charitable interpretation of the workings of the program, finding that it has had some impact of some sort. In addition, our proposed approach will provide deeper understanding of both the social problem involved and the treatments that would be effective. Indeed, one of the possible outcomes of the use of this approach is a closer link between program designers and evaluators at the point of program formulation and design.

Justifying the Importance of Knowledge or Theory-Inferred Outcome Variables

But, some might ask, since program people are only concerned with the official intended effect variables, why bother to include additional inferred effect variables?

As argued in the previous section, it is unlikely that outcomes inferred from theory or knowledge are less important than the official intended outcomes. Official intended effect variables, derived from program goals, are more apt to be desirable, but not necessarily attainable, outcomes. Social science theory and knowledge can tell the evaluator about possible desirable and attainable outcomes, ones which may be more useful in ameliorating the underlying social problem.

In addition, there is the following argument: the ultimate source of program goals are the concerns that agitate policy makers and the public. However, policy makers and public concerns can be expected to change over time. Indeed, it happens often enough that policy makers and the public may lose interest in some goals very soon after a program is created.

Thus at one time, one aspect of a social problem may be blamed as a crucial ill and a program designed to fight that particular aspect. At some other time, another facet of a social problem may be blamed as a crucial ill and another program is created urgently to change that facet. For example, at one time, improving school facilities and teacher quality were regarded as major goals in programs that were intended to improve the learning of disadvantaged students, while increasing parental concern was considered

a major goal at a later time. And further formulations may focus on drop-out rates, achievement motivations, self-confidence, cognitive and affective development, math and reading skills, etc.

Given this condition, it makes a lot of sense to use a multi-goal, theory-driven approach instead of the official-goal-fixed approach to program evaluation, since the new goals might often be found among those effects that social science theory and knowledge indicate as inferred outcomes.

Indeed, these considerations also constitute an argument for including as many sensible outcomes as possible thereby increasing the chances that the goals of tomorrow will be included as well as the goals of today. A case can be made that the less we know about the social problem to which a program is directed and the less we know about the potential workings of a treatment, the greater the number of potential outcomes should be sought for in evaluation.

Sources of Multi-Goal, Theory-Driven Outcome Variables

Not all program effects are worth examining. Some effects are trivial, arising from the very existence of the program itself and are present whether or not the program is at all effective in any meaningful sense. Thus all programs that use up resources, also redistribute resources. Program personnel are hired and hence the program has some effects on the incomes of some persons. Facilities may be rented or purchased. Suppliers provide program needs (e.g., textbooks, medical supplies, etc.). These effects are trivial, of course, from the perspective of ameliorating the social problem to which the program is directed. They may not be trivial to program personnel or to the conglomeration of interest groups that arise around a program.

The program effects of interest from an evaluation viewpoint are those which involve the target population and reflect some alteration in the state of the social problem to which the program is addressed. Clearly, one of the important issues in the application of the multi-goal, theory-driven evaluation approach is how to generate potential outcomes that are not specified in the program goals, as held by program administrators. As we have emphasized, such potential inferred outcome variables are to be derived from social science theory and knowledge. There are some general rules which may be formulated, as follows:

1. THE VARIETY OF SOCIAL SCIENCE THEORIES AND KNOWLEDGE BASES FOR PREDICTING OR FORECASTING PROGRAM TREATMENT EFFECTS

Beyond doubt there is a shortage of grounded theories and knowledge in the social sciences. In addition, the existing theories and knowledge often yield inconsistent and competitive predictions or forecasts about what program treatment effects will be. In the current development of social science theory, it is difficult to ascertain a priori which predictions or forecasts of social treatment effects are definitely superior to others. However, there is no reason why contradictory or competitive outcomes should not be included in the set of potential outcomes to be tested in an evaluation. Indeed, the case can be made, as indicated earlier, that in such circumstances the worth of an evaluation is enhanced by the inclusion of a larger number of potential outcomes. Certainly both social science and policy are better served by the ability of the completed evaluation to decide among competing understandings of the social problems or treatments involved. Hence, all the outcomes deemed possible by social science theory and knowledge should constitute the pool out of which outcomes are to be selected for evaluation testing.

To illustrate our viewpoint, we offer two examples. In the Negative Income Tax (NIT) experiments, a possible effect of income support payments predicted by economic work-leisure tradeoff theory was that NIT would induce men in low income families to quit work or reduce their hours of work (Green). While there is no sociological theory of work directly related to NIT, there are several sociological studies which indicate that work is regarded by people as a source of self-identity and social status (Friedmann and Havinghurst; Goodwin; Morse and Weiss). Hence some sociological knowledge predicted that NIT would not result in work disincentive, since men work to achieve status as well as income.

Furthermore, there are many sociological and social psychological studies which relate income to a variety of social psychological variables and suggest possible outcome variables in NIT experiments (Middleton and Allen; Rossi and Lyall): income relates to self-esteem (Heiss and Owens; Kaplan; Yancey et al.), to feelings of well-being (Bradburn; Bradburn and Caplovitz), to psychological disorder (Dohrenwend and Dohrenwend; Warheit et al.), to anomie (Bell; Srole), to alienation (Bullough; Middleton), to political efficacy (Campbell et al.; Form), etc.

In addition, economic and sociological literature concerning consumption behavior suggested some possible changes in household consumption patterns in areas such as health, education, housing, etc. (Baumol). Malthusian theory suggested a possible outcome would be to encourage recipients of payments to have more children. Finally, since only intact families qualify for payments under NIT, it was reasonable to expect that NIT experiment might increase marital stability.

The second example is the evaluation of the impact of highway construction on social and economic change. Location theory suggested that highway location determines transportation costs which, in turn, affect

the price, supply, and demand for various commodities, locations, and the extent and frequency of human interaction (Isard; Winfrey and Zellner). Some of the effect variables derived from this theory are changes in retail sales, changes in land values, changes in land uses, shifts in land and housing prices, changes in residential populations, etc.

In the other social sciences, theory which can be directly or indirectly used in assessing highway impacts was scarce. However, knowledge of sociology and other social sciences provided some hints that to improve a highway in a community likely would cause changes in the following: levels of living, community values and community organizations (Dansereau); neighboring patterns (Burkhardt, a; Ellis); racial composition, age structure, and other population changes (Grier); social mobility (Hill and Frankland; Mclean and Adkins); physical disruption (Cline); relocation problems (Burkhardt, b). Negative effects on aesthetics, scenery, air pollution, noise, etc., are also suggested by an environmentalist perspective as possible effects of highway improvement (Rollier and Erbetta).

2. IMPLICIT PROGRAM MODELS

Although policy makers and program administrators may not have a very explicit conception of how a social program is supposed to work, the social scientist-evaluator may be able to construct a set of alternative models that are implicit in the program and its goals. For example, in the Contract Learning Experiment, the following implicit model may be inferred: because contractors are motivated by the desire to obtain optimum payments, it is to be expected that they would monitor the learning progress of students in their programs and that they would modify procedures at any sign that student progress was lagging. Hence one would infer that a sign that the program was working as contemplated would be whether contractors set up careful monitoring systems to provide information in real time on student progress. Since without such information, contractors could not be expected to be able to modify procedures, the absence of such a monitoring system would be *prima facie* evidence of a flaw in the program. An alternative model for the same program might be that contractors would stress to program operators (i.e., classroom teachers hired by the contractors) that student motivation to learn was critical and hence urge that operators spend a great deal of time on motivating students, providing yet another measure of the effectiveness of the program.

In short, by careful analysis of a program it is usually possible to infer at least one and usually several implicit models of the program. Indeed, if it is not possible to do so, that fact is also a measure of the effectiveness of a program. In other words, if there is no conceivable model, then there is no conceivable program.

3. INTERVENING PROCESSES

An excellent guide to the construction of multi-goal, theory-driven outcome variables is to follow the rule that all program treatments must operate through a set of intervening processes in order to achieve their goals and that these intervening processes can be expected to have outcomes other than those specified in the program administrators' goals. For example, an income maintenance program's ostensible treatment consists of transfer payments. The delightful characteristic of money is that it may be used for so many different purposes. One of the major interests of the income maintenance experiments was to see whether recipients would use the payments to purchase leisure, i.e., reduce their work effort. But recipients could use the payments in a variety of ways, e.g., purchase more housing, better food, more medical care, support additional children, etc. These additional program effects can be seen as potential outcomes of the treatment, as soon as it is realized that treatments are intervening variables that can lead to a wide variety of potential outputs.

Often enough a program is required to be evaluated before sufficient time has elapsed for ultimate program effects to show themselves strongly enough to be detected. Thus while the general goal of a pre-school intervention program may be to improve participants' performance in later life (perhaps in the later years of elementary and high school), evaluation may be required to take place long before such effects can be measured. Under such circumstances the effects that can be effectively measured are those that measure whether intervening processes have been affected. Hence, such evaluations should focus on student characteristics that may be related to subsequent school success, such as self-confidence, interest in reading and so on.

It may take longer than one year for an alteration in police patrol practices to show effects on crime rates, while one year may be long enough for citizens to develop more confidence in the efficacy of the police, an intervening step on the way to reduce crime rates.

As a final example: the ultimately desired goal of a bilingual program may be to increase participants' income but the program is required to be evaluated a year after the program begins. Under these conditions, it may be advisable to include occupational skills and English ability as outcome variables, since a year may not be long enough for the participants to gain more salary immediately but it is long enough to tell whether the program has affected intervening variables such as occupational skills and language ability.

4. SUB-GROUP DIVERSITY OF TARGET POPULATIONS

A program may work for some groups, but not for others. This possibility also suggests ways to search for inferred outcome variables. For example, overall, the results of the New Jersey–Pennsylvania Negative Income Tax Experiment showed little work incentive or disincentive. However, when the target population was disaggregated, it was found that blacks displayed work incentive effects, whites showed work disincentive effects, and Puerto Ricans showed no effects (Rossi and Lyall). According to Cicirelli, Head Start showed little effect in national total sample, but when subgroups were analyzed, he found the program had affected students in the center of the Southeastern region, or in core cities, which are mainly composed of blacks.

5. GENERAL CHARACTERISTICS OF THE DELIVERY SYSTEM OF A PROGRAM

(1) *Mode of Delivery*

Almost any program needs a delivery system to implement the program. In some cases, the mode of delivery itself will cause some effects which initially were not intended as program goals. For example, an administrator while implementing treatments may add certain administrative conveniences which have nothing to do with the intended treatment and may actually subvert that treatment, or he may interpret the policy and implementation differently from the original plan.

An appropriate illustration is provided by the TARP experiment (Rossi et al.), in which the mode of providing transitional aid funds to ex-felons was the unemployment insurance systems of each of the two states. While these systems were very good at delivering checks and certifying eligibility, ordinary operating procedures led to the strengthening of work disincentives offered by the payments. Unemployment benefit payments are ordinarily administered reducing benefits a dollar for each dollar earned, a practice which makes payments contingent on unemployment and hence enhances the work disincentive effect that such transfer payments would ordinarily be expected to have.

Indeed, careful attention to the delivery system and its mode of proceeding is especially important in human services programs. For example, a program of group therapy for prisoners may be a useful program, but its efficacy may be completely undermined if the group therapists are also prison guards. Or, almost any program can be expected to work when its delivery personnel are highly motivated—as is often the case in pilot programs—but when placed in the hands of an old line bureaucratic agency whose personnel are not as highly motivated the program treatment may be considerably less than optimally implemented. The mode of delivery may also add treatments to those intended. For example, in

income maintenance experiments, treatments must be considered to be both the transfer payments and the accompanying system of earnings reports, audits, and other administrative procedures. Hence a fertile source of additional outcome variables to test for in the evaluation of social programs are the intended and unintended effects of modes of delivery.

A fully developed, theoretically based social engineering would not only have an adequate theoretical understanding of social program but also of treatments. Indeed, the weakest point of existing social science theory and knowledge is precisely that we know so little about treatments and treatment delivery systems. For example, we understand crime better than we understand how to lower crime rates or recidivism.

(2) Participant-Interaction Effects

In some programs, the participants are brought together in a group and interaction among participants facilitated. The intensive interaction among participants and with program people may have some influences on the participants that is over and above that intended. Friendship ties may develop and participants may feel less alienated than before, or they may exchange job information which may lead them to be employed or at least become more motivated to look for a job.

A negative example is that during a prison rehabilitation program an inexperienced thief may learn more sophisticated techniques from experienced inmates, thus becoming more adept at stealing.

6. THE INTERACTION BETWEEN THE PROGRAM AND THE ENVIRONMENTAL SYSTEM

A social program is not carried out within a vacuum, rather, it is ordinarily carried out within social systems—families, neighborhoods, schools, prisons, hospitals, communities, etc. It is possible that either program treatments or program outcomes would interact with social systems and, consequently, additional effects are created. These system effects may either be important in their own right or they may further influence the program outcomes. The interaction between program treatments, outcomes, and its systems environments should also be a potential source for evaluators to examine the inferred outcome variables.

One example of an interaction between program treatments and its environmental system is the Performance Contracting Experiment. It was reported that the school teachers and principals were more willing to try educational innovations after the program was over (Carpenter-Huffman et al.). Another example is the Head Start Program, where it was noted that the disadvantaged students' mothers become more concerned and involved in their childrens' school life and achievement.

An example of interaction between the program outcomes and its

environmental system is a successful halfway house program for prisoners that may change local residents' perception of the prisoners on probation.

Conclusions

Our advocacy of the multi-goal, theory-driven approach to evaluation means that *evaluators should strive to provide information concerning what a program can do, as well as what it cannot do.* Compared to the official-goal-fixed approach, the advantages of applying a multi-goal, theory-driven approach to current program evaluation can be summarized as follows.

1. *A multi-goal, theory-driven approach provides greater opportunities for discerning some non-zero program effects.*

This new approach, by encouraging evaluators to include as many inferred outcome variables as they can according to the situation at hand, increases the possibility of finding some non-trivial program effects.

2. *A Multi-goal, theory-driven approach provides more information for program administrators or policy makers to make better decisions.*

Good decisions can be made best when there is sufficient relevant information. Our approach can provide the decision-makers with information on a wide range of potential program effects, including those related to the nominal goals of the program.

3. *The multi-goal, theory-driven approach is able to contribute to the development of social science theory.*

In form, our approach is similar to basic research. Evaluators using our approach are theory builders who try to develop models that enable one to understand the relations between program variables and outcomes. In addition, since evaluations are usually carried out in the field with more rigorous research methods such as experimental and quasi-experimental designs, the multi-goal, theory-driven evaluation is more likely to provide adequate knowledge of causal relationships among variables, information that is crucial for theory building in any discipline. Using this approach, program evaluation is a challenge to any competent social scientist, and evaluation researchers can contribute to the development of social science theory perhaps as much as any social scientist in basic research.

4. *A multi-goal, theory-driven approach will lead to greater efficiency in the long run in the use of resources for social reform.*

In the short run, our approach may be more expensive than conventional approaches, since it requires the testing of more outcomes. But long-run efficiency may be greater given the savings to be enjoyed from the resulting greater level of program efficiency.

5. *Administrator resistance to evaluation may be less in the multi-goal, theory-driven approach.*

One of the more important sources of administrator resistance to evaluation is that administrators are afraid that evaluators fail to measure the full range of their activities. This fear is not groundless: The measured goals of a broad-aim program can be severely narrowed in the conventional approach to a few measurable outcomes. In contrast, evaluators using our approach, instead of narrowing attention would enlarge the number of outcomes and hence would increase the possibility of capturing program effects. This characteristic of our proposed approach may help to alleviate some administrator anxieties.

Note

1. An outstanding set of exceptions have been the randomized experiments designed by economists to test out the effects of transfer payments of one sort or another, as e.g., The Experimental Housing Allowance Program sponsored by HUD, and the income maintenance experiments started on OEO and continuing under HEW. We will see that the reasons for these exceptions lie in the integration of economic theory into the experimental designs and into the analyses of treatment effects.

References

- Angrist, S. S. 1975. "Evaluation Research—Possibilities and Limitations." *Journal of Applied Behavioral Science* 11(1):75–96.
- Baumol, W. J. 1977. "An Overview of the Results." In Harold W. Watts and Albert Rees (ed.), *Income-Maintenance Experiment*. Vol. 3. New York: Academic Press.
- Bell, W. 1957. "Anomie, Social Isolation, and the Class Structure." *Sociometry* 20(June):105–16.
- Bradburn, N. M. 1969. *The Structure of Psychological Well-Being*. Chicago: Aldine.
- Bradburn, N. M., and D. Caplovitz. 1965. *Reports on Happiness*. Chicago: Aldine.
- Bullough, B. 1967. "Alienation in the Ghetto." *American Journal of Sociology* 72(March):469–78.
- Burkhardt, Joe F. a:1971. *Impact of Highways on Urban Neighborhoods: A Model of Social Change*. Highway Research Record 356. Washington: National Academy of Sciences, Highway Research Board.
- . b:1973. *Community Reactions to Anticipated Highways Fears and Actual Effects*. Highway Research Record 470. Washington: National Academy of Sciences, Highway Research Board.
- Campbell, A., P. E. Converse, W. E. Miller, and D. R. Stokes. 1960. *The American Voter*. New York: Wiley.
- Carpenter-Huffman, P., G. R. Hall, and G. C. Sumner. 1964. *Change in Education: Insight from Performance Contracting*. Cambridge: Ballinger.
- Ciccirelli, V. 1969. "The impact of Head Start: Executive Summary." Reprinted in Francis G. Caro (ed.), *Reading in Evaluation Research*. 1st ed. New York: Russell Sage Foundation, 1971.
- Cline, Marvin G. 1963. *Urban Freeways and Social Structure: Some Problems and Proposals*. Highway Research Record 2. Washington: National Academy of Sciences, Highway Research Board.
- Dansereau, H. K. 1975. *Five Years of Highway Research: A Sociological Perspective*.

- Highway Research Record 75. Washington: National Academy of Sciences, Highway Research Board.
- Deutscher, I. 1977. "Toward Avoiding the Goal Trap in Evaluation Research." In Francis G. Caro (ed.), *Reading in Evaluation Research*. 2d ed. New York: Russell Sage Foundation.
- Dohrenwend, B. P., and B. C. Dohrenwend. 1969. *Social Status and Psychological Disorder: A Causal Inquiry*. New York: Wiley.
- Ellis, Raymond. 1968. *Toward Measurement of the Community Consequences of Urban Freeways*. Highway Research Record 229. Washington: National Academy of Sciences, Highway Research Board.
- Form, W. H. 1973. "Auto Works and Their Machines: A Study of Work, Factory, and Job Satisfaction in Four Countries." *Social Forces* 52(September):1-15.
- Friedmann, E. A., and R. J. Havinghurst (eds). 1954. *The Meaning of Work and Retirement*. Chicago: University of Chicago Press.
- Gilbert, John P., Richard J. Light, and Frederick Mosteller. 1975. "Assessing Social Innovations: An Experimental Base for Policy." In Caro Bennett and Arthur A. Lumsdaine (eds.), *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs*. New York: Academic Press.
- Goodwin, L. 1972. *Do the Poor Want to Work? A Social-Psychological Study of Work Orientation*. Washington, D.C.: The Brookings Institution.
- Green, C. 1968. "Negative Taxes and Monetary Incentives to Work: The Static Theory." *Journal of Human Resources* 3(Summer):280-88.
- Grier, George W. 1970. *Social Impact Analysis of an Urban Freeway System*. Highway Research Record 305. Washington: National Academy of Sciences, Highway Research Board.
- Heiss, J., and S. Owens. 1972. "Self-Evaluations of Blacks and Whites." *American Journal of Sociology* 78(September):360-70.
- Hill, Stuart L., and Bamford Frankland. 1967. *Mobility as a Measure of Neighborhood*. Highway Research Record 187. Washington: National Academy of Sciences, Highway Research Board.
- Isard, Walter. 1956. *Location and Space Economy*. Cambridge: M.I.T. Press.
- Kaplan, H. R. 1971. "Social Class and Self-Derogation: A Conditional Relationship." *Sociometry* 34(March):41-64.
- McLean, Edward L., and William G. Adkins. 1971. *Freeway Effects on Residential Mobility in Metropolitan Neighborhoods*. Highway Research Report 356. Washington: National Academy of Sciences, Highway Research Board.
- Malthu, Thomas R. 1963. *An Essay on the Principle of Population*. Homewood, Il.: Irwin.
- Middleton, R. 1963. "Alienation, Race and Education." *American Sociological Review* 28(December):973-77.
- Middleton, R., and V. L. Allen. 1977. "Social Psychological Effects." In Harold Watts and Albert Rees (Eds.), *The New Jersey Income Maintenance Experiment*. Vol. 3. New York: Academic Press.
- Morse, N. C., and R. S. Weiss. 1955. "The Function and Meaning of Work and the Job." *American Sociological Review* 20(April):191-98.
- Riecken, Henry W., and Robert F. Boruch. 1974. *Social Experimentation: A Method for Planning and Evaluating Social Intervention*. New York: Academic Press.
- Rollier, M., and M. A. Erbetta. 1976. "Environmental Impact on Highways." *Ekistics* 42:37-21.
- Rossi, Peter H. 1978. "Issues in the Evaluation of Human Services Delivery." *Evaluation Quarterly* 2(November):573-99.

- Rossi, Peter H., and Katharine Lyall. 1976. *Reforming Public Welfare: A Critique of the Negative Income Tax Experiment*. New York: Russell Sage Foundation.
- Rossi, Peter H., Richard A. Berk, and Kenneth J. Lenihan. 1980. *Money, Work, and Crime*. New York: Academic Press.
- Scriven, M. 1972. "Pros and Cons About Goal-Free Evaluation." *Evaluation Comment* 3:1-4.
- Srole, L. J. 1956. "Social Integration and Certain Corollaries: An Exploratory Study." *American Sociological Review* 21(December):709-16.
- Warheit, G. J., C. E. Holzer, III, and J. J. Schwab. 1973. "An Analyses of Social Class and Racial Differences in Depressive Symptomatology: A Community Study." *Journal of Health and Social Behavior* 14:291-99.
- Weiss, R. S., and M. Rein. 1969. "The Evaluation of Broad-Aim Programs: A Cautionary Case and a Morale." *Annals of the American Academy of Political and Social Science* 385(September):133-42.
- Winfrey, Robley, and Carl Zellner. 1971. *Summary and Evaluation of Economic Consequences of Highway Improvements*. National Cooperative Highway Research Program Report 122. Washington: National Academy of Sciences, Highway Research Board.
- Wholey, J. et al. 1975. "Evaluation: Where is it Really Needed?" *Evaluation Magazine* 2(2):89-93.
- Yancey, W. L., L. Rigsby, and J. D. McCarthy. 1972. "Social Position and Self-Evaluation: The Relative Importance of Race." *American Journal of Sociology* 78(September):338-59.