CONCEPTUAL APPROACHES TO ASSESSING IMPACTS OF LARGE-SCALE INTERVENTION PROGRAMS

Howard E. Freeman, Brandeis University

Dissatisfaction with the social order and zealous efforts at community change have characterized the personal and academic lives of social scientists since their emergence as an identifiable group on the American scene.[1] In many ways, of course, the several disciplines and the persons that held membership in them have changed markedly over the last several decades: much to the despair of some, the influence of visionary clergymen, guilt-ridden do-gooders, and political radicals--dedicated to projecting their own humanitarian views in the guise of scientific inquiry--has pretty well diminished.[2]

Today the social scientists' role in the modification of community life and the amelioration of social pathologies is a much different one. He puts forth theories on which action programs may be based, he serves as expert and consultant to policy-makers and he uses his research repertoire to guide program development. Admittedly, much of social science activity is directed at understanding "basic" processes, but, whether by intent or unwittingly, he serves as an agent of social change; and, if one is willing to extrapolate from shifts in occupational settings, it appears that there is a growing movement of research persons who know full-well of the social-change potential of their work.[3] Certainly there are outstanding examples of influence: the work of Stouffer and his associates on military problems, the studies of learning psychologists on educational practices, the manifesto of Clark and other social scientists in connection with the Supreme Court's integration decision, and most recently the document of Ohlin and Cloward on delinquency programs.

During the past 15 years, with the increased emphasis--particularly at federal level--on demonstration programs, there has been much concern with the assessment of therapeutic and rehabilitation efforts and thus the developments of a sub-specialty most typically referred to as evaluation research. Virtually all of the demonstration programs supported by federal funds in the health and welfare field and many of the projects sponsored by philanthropic foundations include a requirement that the worth of the effort be assessed.

For the most part, however, the requirement on evaluation has remained a formality; granting agencies have tended to overlook it in their frenzy to implement programs intuitively believed worthwile, statements and often elaborate designs for evaluation in demonstration--research programs have been included in proposals as a ritual with full knowledge that the commitment would not be met, and researchers have, on occasion, found it expeditious to accept evaluation assignments and then redirect the resources to another type of study.

It is only fair to acknowledge the minimal contribution that we have made to program development and social policy through evaluation studies. To some extent the limited impact of previous evaluation research is related to difficulties of successfully implementing and conducting experimental investigations and to barriers put forth by practitioners. There is no need to underscore the difficulties of undertaking research when the cooperation of practitioners and flexibility on their part is necessary for the development and implementation of an adequate design; conflict between clinician and scientist pervades all fields and the difficulties that medical researchers have in undertaking experiments with human subjects are minimal in comparison with evaluation efforts in the community.[4] Also, of course, many social scientists engaged in evaluation studies regard them as a dilittante activity and their interest in such work centers about partially testing a theory that they are concerned with or because it provides publications and sometimes economic affluence. But the major reason I contend that accounts for the minimal impact of evaluation studies on programs and policy is the sheer infrequency that adequately conceived efforts have in fact been undertaken. Whatever be the explanation, certainly it is difficult to point to many instances in which programs actually have been expanded or terminated because of evaluation findings.

---

[1] Howard Odum, American Sociology, New York: Longmans Green, 1951.

[2] Maurice Stein, Sociology on Trial, New York: Prentice-Hall, Inc., 1963.

[3] E. Sibley, Education of Sociologists in the United States, New York: Russell Sage Foundation, 1963.

[4] R. C. Fox, Experiment Perilous: Physicians and Patients Facing the Unknown, New York: The Free Press of Glencoe, Inc., 1959.

We no longer have the opportunity, however, to learn new methodological wrinkles or to develop slowly and carefully a strategy for rendering the results of evaluation studies into a potent force in the determination of action programs and social policy. Suddenly we have a mandate to participate in massive social change, via community-wide efforts projected to restructure health and welfare activities and to reorient the efforts of practitioners. Despite the failure to work out fully methods and most important a strategy to influence policy on small-scale action programs, we now have been thrust into a predominant role in these massive efforts, one designed to have an impact on virtually all community members and indeed on the very social order. It is simply not possible to retreat from this assignment, any more than it is for all physicists to avoid participation in the development and improvement of destructive devices.

But our position is a dangerous one. Although many individuals, for a variety of reasons, have decried so-called centralized programs of planned change and have expressed alarm over their control by public bodies and large foundations, apparently this is the direction that health and welfare activities are going to take; and, if one may regard the recent election in this sense, certainly there is an overwhelming mandate for these efforts to continue.[5] Perhaps those of us located in professional schools or employed directly by community-based programs are most sensitive to the stakes, but it is obvious that the comprehensive and massive character of projects sponsored by the President's Committee on Juvenile Delinquency and Youth Crime, the Ford Foundation, and now the Office of Economic Opportunity are likely to rock the very foundations of our social system.

Unfortunately, although we have improved at least to some extent our technical repertoire, we have not been successful in developing and explicating a set of conditions that must be met in order for our work to have social policy potential. In a previous paper I suggested that because of a concern not to lose our identification with our academic disciplines and not to sacrifice objectivity, we have not considered how studies need to be developed in order that they have an impact on program and policy persons and have failed to recognize the vested interests of the various groups involved in program and policy-making.[6] Although I will touch on a number of issues here, I would like to focus on two points: the need to define our role appropriately given the environment in which we are being called upon to work and the necessity to organize data and findings so that their potential utility in program and social-policy development is maximized.

The Research Environment

Since most of us have at one point or another been involved with large bureaucracies operating on a continual crash basis, certain rather obvious observations can be made most briefly. It is important to point out that dependence upon the legislature branch or the whims of foundations for funds and the necessity to involve and obtain the cooperation of politically and ideologically antagonistic parties in local communities have led up to now and I am sure will continue to evoke a considerable degree of disorder in most of the massive programs. The development of adequate staffs, personnel policies, and long range planning by community-based mass programs is difficult, some maintain almost impossible given the condition of being affluent one minute and poverty-stricken the next and given the fleeting support of the various political forces involved. The shape, size and goals of programs appear to change from day-to-day, and one of the difficulties of evaluation research in these settings stems from the high degree of organizational and interorganizational chaos.

Even in those efforts in which the overall objectives remain relatively stable, the number of specific programs is large and their goals diverse. The evaluation design needs to be developed in terms of a series of staged inputs and outputs, the situation is much too complex to fit the classical independent-dependent variable model. Further the image of the researcher who remains outside the environment and evaluates what others are doing in no way squares with the reality of his engagement in these programs. It is clear that the researcher is involved in a situation in which he must lock himself into the environment, not only because he has a background that can be exploited by persons designing programs, but because otherwise he cannot accomplish his evaluation task. Unless he participates, indeed leads

---

[5] John R. Seeley, "Central Planning: Prologue to a Critique," in Robert Morris (ed.), Centrally Planned Change: Prospects and Concepts, New York: National Association of Social Workers, 1964, pp. 41-68.

[6] Howard E. Freeman, "The Strategy of Social Policy Research," in The Social Welfare Forum 1963, New York: Columbia University Press, 1963, pp. 143-156.

the dialogue and bargaining required for the identification of goals, for description of input-output variables, and for the elaboration of a rationale that specifies the relationship between input variables and goals, these tasks are likely to remain undone. Once formulated he must continue to remain within the environment, like a snarling watchdog ready to fight alterations in program and procedures that could render his evaluation efforts useless.

It is only fair and from my view unfortunate to note that he can expect little help or guidance from the funding groups in these tasks. In part this is related to the lack of structured expectations of outcome on the part of these groups, but also because of an effort to maintain as non-directed a posture as possible in the light of accusations of authoritarian control. The various President's Committee on Delinquency projects illustrate this point well: from city to city, though the legislation directs attention to the reduction of youth crime and the amelioration or related problems, considerable, no great latitude, has been allowed in individual cities not only in program development but in evaluation design. Thus, not only is there variations in whether one is concerned with area-crime rates, the police contacts of individual youths, or the reduction of deviant though not necessarily illegal behavior, but some cities apparently have not felt a need to be very concerned with any measure of this sort. Unless the situation changes, the researcher is naive to expect that sanctions from above are going to provide him with much support in the specification of objectives, the identification of goals of sub-programs or the outlining of the theoretical links between the goals of sub-programs and overall objectives. The researcher has three choices: he can follow Hyman's recommendation and try to guess the intermediate and over-all goals, and later be told that the ones he selected were not relevant at all; he can insist that program persons provide them in which case he should bring lots of novels to the office; or he can participate or rather take a major responsibility for the development of the action framework.[7] In the companion paper, the impact model of a particular program will be discussed in some detail and this point amplified--but I do maintain that there is little likelihood of developing evaluation designs for these massive programs by either second-guessing the action people or by insisting upon their coming up with an appropriate and explicit flow-chart.[8] Indeed, if the researcher is going to act responsibly as an agent of social change through his evaluation research, I contend it is mandatory for him to engage himself in program development.

I do feel, however, that the task would be much easier if the sponsors of these massive programs would establish and enforce a requirement that the necessary specifications be part of any application and renewal of applications and that they exercised sanctions to prevent slippage. Furthermore, that the sponsors provide a minimal set of outcome variables--uniform measurement would be most valuable for long-range program planning. It is most difficult, indeed probably impossible, to compare the various delinquency prevention efforts of the last three years, the various mental health reorganization attempts over the past ten years and, unless there are marked changes in policy, only limited likelihood of making city-to-city comparisons in the economic, educational and occupational rehabilitation programs now underway as part of the poverty package. Given the lack of structured directions by the government and foundation granting programs, and the lack of commitment to evaluation research on the part of many practitioners on the local level, it is not easy to manipulate the environment so the researcher can undertake his task.

I must acknowledge that the researcher has not always participated in these evaluation studies enthusiastically and with a full sense of commitment; to argue that the problems of evaluation research are solely due to the actions of others is as ludicrous as the general who maintained that the high V.D. rate among his troops was due to the promiscuity of the civilian population. Participation within the action environment obligates the researcher to bring to bear his substantive knowledge in the design of programs and to be a positive influence in their development and to recommend and condemn program plans or at least forcefully report and interpret findings from other research that have a bearing on program development. This we often

---

[7] Herbert Hyman, Applications of Methods of Evaluation for Studies of Encampment for Citizenship, California: University of California Press, 1962.

[8] Clarence C. Sherwood, "Methodological, Measurement and Social Action Considerations Related to the Assessment of Large-Scale Demonstration Programs," read at the annual meeting of the American Statistical Association, Chicago, Illinois, December 29, 1964.

fail to do. If we did exercise our responsibility, we probably would have built into these massive efforts more attempts to use physical means such as brighter street lights to prevent delinquency and have exerted more pressure for cohesive programs such as forced literacy training as a condition of probation and parole in contrast with increased numbers of therapeutic communities and the burgeoning of street-worker projects.[9] The researcher if he is to be part of the action environment must share responsibility not only for the niceties of the evaluation design but for the general over-all frame of the agency.

Most to the point, however, is the researcher's reluctance in many cases to adopt a client perception of his work: many of us in our academic endeavors appropriately behave like small businessmen with a monopoly in a particular neighborhood. An entreprenuial style is not consistent with the research needs in these mass programs and our academic colleagues cannot be the only reference groups for our actions.[10] I contend that research in these mass programs must be regarded as client-sponsored investigations, and variables included in these studies need to take into account the client, i.e., the policy-makers who provide the means and support for the program, not the editors of our journals nor--and I am sure some of my colleagues will not like the remark--the recipient populations. The soap manufacturer is interested in how many boxes of soap he sells and not in whether infant children like the color of the box. In the same way, there is no reason for the policy-maker to be concerned with esoteric measures that might occupy some novel place in an academic field and thus, for example, whoever is paying for a reading program needs and deserves to be told whether it improves kids' reading and not whether it reduces hand tremors. The way it seems to me to approach this latter problem is to have available several conceptions of evaluation and it is to this matter that I would like to turn.

Conceptions of Evaluation

Most of our methodology in evaluation studies stems, of course, from much more highly simplified situations; experimental designs that have worked in agriculture, in the life sciences and in rat psychology. We must be concerned with efficacy of outcome and experimental studies are required in order to determine efficacy. Evaluation research efforts must seek to approximate the experimental model as much as possible--we do not do so often enough and some of the so-called evaluation designs of the current mass programs have completely foregone an experimental or quasi-experimental approach. Admittedly, there is a limit to the extent controlled experiments can be conducted in these programs but nevertheless it is possible in most instances to make use of at least rudimentary or quasi-designs. In the companion paper some of the barriers to experimental research in community programs will be discussed but it seems to me obvious that to do our job it is necessary, be it through randomization or statistical procedures, to approximate the conditions of before-after and/or pre- post test designs.

The situation becomes exceedingly complicated, however, given the need to evaluate a linked input-output system rather than being able to examine specific independent-dependent variable situations. But the kinds of massive efforts going on are of a linked input-output type and it is necessary to measure efficacy of each of the specific programs, the interactions between programs, and to be able to tie together by means of relational analysis the impact that changes due to sub-programs have on the over-all program objectives. For example, an educational program may be designed to improve reading and this must be assessed but if the over-all objective of the community project is to reduce school drop-outs, the relationship between reading improvement and drop-outs must also be demonstrated.

Of the many problems we are up against in the utilization of experimental models, the linking issue seems to me to be the most difficult. Among other things, we are too willing to make use of reflectors of program change rather than program change itself, such as employing shifts in attitudes toward Negroes when the program is concerned with reducing discrimination. Most of us are aware of the limited correlations often found between attitudes and behavior, but as a recent paper points out, the situation may be worse than that: reanalysis of several studies suggest that changes in attitudes may be inversely correlated with changes in behavior. Thus, if one may extrapolate, reducing prejudice may indeed lead to increasing discrimination.[11] Use of attitudinal reflectors may therefore render impossible the linking process.

---

[9]Admittedly, the evidence about the latter two approaches is fragmentary but nevertheless hardly in the direction to encourage the current expansion efforts. See Charles Perrow, "Hospitals: Goals, Structure and Technology," in James March (ed.), Handbook of Organizations, New York: Rand McNally, 1964 and Walter B. Miller, "The Impact of a Total-Community Delinquency Control Project," Social Problems, 10 (Fally, 1962), pp. 168-191.

[10]Howard E. Freeman, loc. cit.

[11]Leon Festinger, "Behavioral Support for Opinion Change," Public Opinion Quarterly, 28 (Fall, 1964), pp. 404-417.

Another dilemma comes about because of our reluctance to keep program efforts within the bounds that permit the maintaining of adequate control groups. In many instances we fail to keep the sub-programs at a level that makes possible the establishment of control groups and thus community programs for previously institutionalized delinquents, nursery-school programs for children of working mothers, and home-maker efforts for the aged are often implemented in such ways that though there are administrative opportunities for randomization nevertheless do not occur. I believe Dr. Sherwood will elaborate this point but I must note that if the "problem" group is so small that some subjects cannot be reserved for a control group, one should question its appropriateness in these mass programs.

Given the size of community efforts under this poverty program, assessing the efficacy of each sub-program in every city is pretty well impossible. Even assuming the availability of research funds, the problem of obtaining necessary professional manpower renders this an unworkable task. Consequently it probably is necessary to sample programs in various cities and this raises knotty problems because of the already-made observations of the linked input-output character of these programs. Sampling must be attempted in terms of the selection of linked programs and the sampling unit needs to be a sub-system of linked programs, analytically if not actually distinct. For example, if one of the goals of a day-care program is to free unmarried mothers so they may receive work-skill training so they may be eligible for employment counseling and training, this "sub-system" of programs must constitute the sampling unit. In order to sample such linked programs, however, it is necessary to have explicit statements of the goals and linkages of the various parts of the community-wide efforts and the necessity to sample speaks too for the need of well-formulated conceptual frameworks for the current efforts.

I say, however, that we have no alternative but experimental evaluation. Should we demand less in terms of the treatment of community problems than we call for in the provision of medical care for ourselves or our pets? Despite the problems of limited sampling and of validity and reliability, many of us seek out Consumer Reports before making major purchases and a few of us even query our physicians about the efficacy of his intended therapies. We reject notions of "intuitive reasonableness" and "impressionistic worth" and seek out comparative assessments before purchasing major consumer items and we have the responsibility to stand for such evaluations in these mass programs as well.

At the present time even the most basic aspects of these efforts are open to question. Many of the mass effects, for example, are heavily committed to community organization programs and to the stimulation of expressive actions on the part of the so-called deprived populations. These programs have as in the case of New York City's Mobilization for Youth been a major source of controversy and yet, despite the resources expended and the conflict occasioned by them, at present they cannot be condemned or condoned in terms of objective evidence.[12] It is possible to mass opinions pro and con but such major issues cannot be settled given the current state of things though it is thirty years ago that community experiments were attempted by a social scientist in Syria.[13]

Let me move on to the second requirement of evaluation research, namely, that of accountability. By accountability I mean evidence first that there is indeed a target population that can be dealt with by means of a program; second, that this population is important either because of its size or the intensity of pathology; and third, that the project program for the target population is undertaken actually with them.

It is not enough to evaluate efficacy--the outcomes of programs--the massive efforts now underway need to be evaluated in terms of accountability as well. While one might be accused of being inhuman for saying it, given the needs, there is little excuse for sanctioning action programs that effect insignificant portions of the population. One of the aspects of accountability is the estimation of the incidence and prevalence of problems. Oftentimes, I am afraid, programs are developed to deal with problems that exist in the minds of practitioners or because of stereotypes held by the public.

---

[12] Roland L. Warren, "The Impact of New Designs of Community Organizations," paper presented at the annual meetings of the National Social Welfare Assembly, November 3, 1964, New York City.

[13] Stuart C. Dodd, A Controlled Experiment on Rural Hygiene in Syria, Beirut, Lebanon Republic; American Press, 1934.

If I may cite one illustration, it is the drug-addiction problem: despite newspaper and public alarm, the incidence in many urban centers is so low that on accountability grounds these efforts hardly merit the attention of so many or the utilization of extensive research resources to evaluate them. If small-size programs use up all the potential clients, then it appears to me only under very unusual circumstances may the researcher be justified in collaborating in their evaluation or even attempting to do so. If the programs are of a large-scale type, than the denial of services or the provision of "ordinary" treatment to a few for control purposes and subsequent estimation of worth is entirely necessary.

Accountability, however, has to do with more than the number of clients served and the size of the potential aggregate of them. Evaluation researchers, in addition to a responsibility for determining efficacy, must deal with the implementation of the prescribed process. In many instances we have engaged in outcome studies without having any knowledge of whether or not what program people maintain is going on actually takes place. This point, I am sure, will be expanded on in the next paper, but it is clear that in many of the sub-programs being implemented as part of these massive efforts--even when evaluation studies of the finest design are accompanying them--we are estimating the utility of programs that never get off the ground; evaluating programs in which volunteers do no more than sign up or week-end educational camping programs in which kids have a good time and do nothing more than play ball or eat marshmallows around the fireside. To say a program fails when it is not truly implemented is indeed misguided, and the evaluation researcher's responsibility here is one of providing evidence and information that permits an accounting of what took place as well as what was the result.

Finally, what we hardly ever worry about, to my knowledge, is efficiency.[14] The various specific programs that are linked together in these massive packages differ extensively in target groups, use of scarce resources and duration. At the risk of being ludicrous, suppose neither individual psychotheray nor group psychotherapy has any impact on the lives of persons but the former costs ten times that of the latter, given such a situation I know where I would put my money. In certain fields of medicine and in certain areas of welfare there is literally no way, given the community's ideological outlook, to cease all treatment even if no efforts are efficacious. But without being too cynical, even when we know this is the case, we refuse to employ a concept of efficiency. Suppose short-term treatment institutions for delinquent offenders do no better than long-term ones, if they are more economical is this not something that the evaluation researcher has a responsibility to take into account?[15]

In terms of all programs, the efficient one is that which yields the greatest per unit change not the one that can be run at the least cost per recipient. What costs the most, takes the longest, and involves the greatest amount of manpower in gross terms may have the greatest net efficiency.[16] Decisions on the continuance of various programs beyond trial--demonstration periods require that we think in these terms. In most evaluation efforts, I would argue we fail to make use of per capita costs, per manpower costs or per time measures.

I would contend that concepts of accountability and efficiency as well as efficacy need to be implemented in order for evaluation research to be properly undertaken. Admittedly, we ought to seek out efficacious programs. But these programs are or at least should be accountable in order for policy and program persons to make rational decisions, and we must also concern ourselves with efficiency of operations.

Concluding Comments

I hope my remarks, though not entirely original, of course, may prove relevant for researchers who have occasion to participate in the evaluation of community-wide programs. The need to become engaged in the action environment, to look at a linked input-output system and to insist on experimental designs, and the necessity to assess efficiency and to recognize the accountability function in evaluation are, to my mind, key points and ones not well-documented in our methods books and not always held to by persons participating in the evaluation of these massive efforts.

---

[14] An illustration of a study that does consider this problem is Julius Jahn and Margaret Bleckner, "Serving the Aged," (Methodological Supplement--Part I), New York: Community Service Society of New York, 1964.

[15] Howard E. Freeman and H. Ashley Weeks, "Analysis of a Program of Treatment of Delinquent Boys," American Journal of Sociology, 62 (July, 1956), pp. 56-61.

[16] Clarence C. Sherwood, "Social Research in New Community Planning Organizations," paper presented at the National Conference of Social Welfare, Cleveland, 1963.

But I would like to feel that I have communicated more than some specific observations--that I have conveyed the potentiality of our role as change agents and the sense of conviction, commitment and responsibility required on our parts. At no other point in time have we had so great an opportunity to have an impact on the social order; if we are to realize our potential within our current stance as social scientists, however, we need more than additional technical innovations. It is an outlook, an ideology, almost a morality if you will, that we must develop in order to function appropriately as agents of social change.