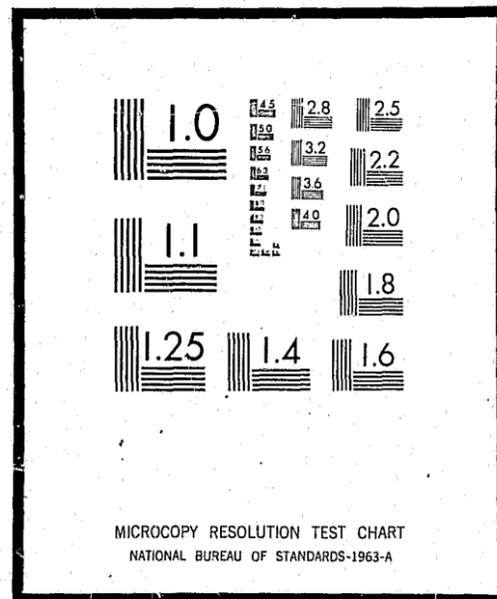# NCJRS

This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.

1.0	145	2.8	2.5
	140	3.2	2.2
		3.6	
1.1		140	2.0
			1.8
1.25	1.4		1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U.S. Department of Justice.

U.S. DEPARTMENT OF JUSTICE
LAW ENFORCEMENT ASSISTANCE ADMINISTRATION
NATIONAL CRIMINAL JUSTICE REFERENCE SERVICE
WASHINGTON, D.C. 20531

Date filmed 12/19/75

---

non loan

THE CONDUCT OF EVALUATIVE RESEARCH OF

FEDERALLY FUNDED SOCIAL ACTION PROGRAMS:

WITH SPECIFIC REFERENCE TO PROGRAMS IN THE

ADMINISTRATION OF JUSTICE

Dorothy F. Berg

Office of Crime Analysis
Office of Criminal Justice Plans and Analysis
Government of the District of Columbia

January 1974

26160 Evaluation

C O N T E N T S

I.   INTRODUCTION

Wider use of program evaluation and analysis of social action and social intervention programs is being urged by legislators, federal agencies, and social scientists as contributions to determination of program efficacy and fiscal responsibility. Along these lines, Senator William V. Roth, Jr. of Delaware (1972:8), commenting on a survey of federal program evaluation practices prepared by his staff, had the following to say about evaluation as it concerned state and local grants-in-aid:

> I am most hopeful that the Federal government will in the future take more interest in encouraging state and local government capacity to manage intergovernmental aid minus Federal requirements. Following upon this concern, in our questionnaire we asked agencies to comment on their efforts to foster evaluative ability among state and local grant recipients. Both executive departments and independent agencies made it clear that almost no programs to support improvements in evaluation and analysis exist. Similarly, almost no functional programs permit the use of money for such purposes.
>
> If we were to help our states and localities develop more capacity for self-criticism, we might be able to eliminate much of the expensive red-tape and bureaucracy now involved in administering Federal domestic assistance. As a consequence some of those at all levels of government who had formerly administered the endless requirements associated with categorical grants might be trained to access the accomplishments of grants-in-aid... (Roth, 1972:8).

In keeping with the need for greater use of evaluation and the stress placed upon evaluation by Senator Roth, the purpose of this paper is to present a synthesis and analysis of the field of evaluative

research which will serve as a guide to staff members in the Office of Criminal Justice Plans and Analysis to: (1) develop and implement a plan for the evaluation of criminal justice programs as required by LEAA and (2) undertake studies to determine the efficacy of the programs funded by this office. These studies should provide decision makers both at the state and federal levels with information on whether program outcomes are equivalent to program goals within budgeted costs. Senator Roth (1972:3) in discussing the needs of the Federal government vis a vis evaluative _research_ has this to say:

> My interest in making sure that the Executive Branch and the Congress have adequate evaluation and analysis to back up their decision making is derived from a desire to find a practical path to true fiscal responsibility. Evaluation and analysis contribute to this end by allowing us to better determine whether programs are accomplishing their intended goals; how these programs could be improved; and what new programs should be undertaken.

State Planning Agency (SPA) programs are financed with federal funds granted by the Law Enforcement Assistance Administration (LEAA) which has established guidelines for program evaluation. These are:

1. Evaluation of 15 percent of all projects funded,

2. Evaluation of 15 percent of the dollar value of all projects funded,

3. Evaluation of all projects within a significant program category.

¶ The framework of this paper will address itself to the following topics: evaluative research and social programing: concepts and issues; issues

and problems in the methodology of measurement and design of evaluative

research; and problems in evaluating intervention programs in social *action*

~~section~~ agencies.

II. EVALUATIVE RESEARCH AND SOCIAL PROGRAMMING: CONCEPTS AND ISSUES

In its commonsense meaning, evaluation is defined as the process of examination and judgment. In government and business, evaluation has taken on several meanings and has been applied in varying ways.

Types of Evaluation

1. <u>Evaluation as evaluative research studies</u> -- This type of evaluation refers to the use of the scientific method for collecting data concerning the degree to which the goals of a social action program are achieved. It is, thus, an attempt to measure by "fact-finding methods that yield evidence that is objective, systematic, and comprehensive" (Hyman and Wright, 1971:185), the amount of change resulting from a social action program. However, in the process of determining how effectively the goals have been accomplished, it is necessary to examine the entire process of social programming beginning with the inception of the program. Thus, analyses are made of input into the program, how the program is carried out (i.e., thruput), whether the goals of the program are worthwhile, what the effects (i.e., output) of the program are, and how efficient the output (in terms of goal fulfillment) is in relation to the input (i.e., in terms of dollars, services and/or personnel time). This type of research is the subject matter of this paper.

2. <u>Evaluation as a process of making judgments of worth</u> -- In this sense administrators [*administration*] evaluate action programs by using their intimate knowledge of the functioning of a particular program as a logical and rational basis for making judgments. Such judgmental evaluation does not require the evaluator to gather empirical evidence to support his conclusions. Evaluation of this kind is useful when a program is small, low cost, and the administrator is very close to all aspects of the program. This type of evaluation is not effective for large scale or long-range programs and does not offer the decision maker a rational basis for retaining or eliminating programs. This is not the type of evaluation which is included in the LEAA guidelines to State Planning Agencies.

3. <u>Evaluation as program analysis</u> -- In this meaning evaluation is accomplished during the planning phase of social programming prior to program implementation. {This type of analysis} consists of a systematic {review} of alternative ways of meeting government objectives. This means that anticipated costs and anticipated effects of alternative programs are compared within a framework of budget ceilings and government objectives. In the process of analyzing program alternatives, cost-benefit analysis is utilized as an important tool but not the only tool. The purpose of program analysis in the planning phase is to optimize resource allocation and to provide information to decision makers concerning the major tradeoffs and implications of the alternatives which have been considered. This type of prior evaluation will be considered in

this paper only insofar as it is relevant to an understanding of the total process of social programming.

Historical Perspective Of Evaluative Research

The earliest evaluative research was conducted in the field of education at the end of the nineteenth century. Between 1920 and 1940, several research studies were carried out to determine effects of social intervention programs in the variety of settings. During the second World War, extensive evaluative research on the effects of films and other forms of mass communication was conducted by a group of sociologists headed by Samuel Stouffer then with the Research Branch of the Information and Education Division of the U.S. Army (Caro, 1971b:5). After World War II, a group of social psychologists, following the lead of Kurt Lewin, engaged in evaluative research which measured "effects of programs designed to change attitudes toward minorities, effects of programs designed to apply group dynamics principles in industry, and the effects of community organization activities on the morale of residents of a housing project" (Caro, 1971b:5). Other evaluative efforts included study of a volunteer work camp and of an innovative program for the treatment of delinquents. During the 1960's when social problems were rediscovered and a sympathetic Federal administration put into effect many social action programs, renewed interest in evaluative research was expressed and carried out on many programs.

Concepts

There are several concepts which will be recurring throughout this paper. Therefore, it would be useful to define them and thus place them in their proper perspective. These concepts are (1) scientific method, (2) goals, (3) program and project, and (4) social action and social intervention.

1. <u>Scientific method</u> -- The purpose of evaluative research as well as research of all types is to answer questions. This is done through the use of the scientific method, i.e., the application of scientific procedures. These procedures increase the likelihood that the data collected will be relevant to the questions asked and will be reliable and valid. In order to be answerable by the scientific method, questions must have one characteristic in common: "They must be such that observation or experimentation in the natural world (including...the behavior of human beings) can provide the needed information" (Selltiz, et.al., 1951:3). Some questions cannot be answered by the scientific method at the present time because procedures have not yet been devised to gather the relevant data. For example, there are no known procedures to devise universally applicable, psychological tests.

2. <u>Goals</u> -- Webster (1971: 358) defines goal as the end toward which effort is directed. The term objective which is similarly defined will be used interchangeably with goals in this paper. Later discussion will center around identification, formulation, and measurement of goals of social intervention programs.

3. Program and project -- These definitions were developed at a Region 2 LEAA Grant Evaluation Conference (May 25, 1971) in which members of OCJP&A

OCJPA participated.

a. "A program is defined as a set of objectives, strategy, and projects to achieve specific program goals" (Report: 1971).[2]

b. "A project is defined as a set of related activities or tasks to achieve the common objective" (Report: 1971) of the program. A project may be one of many to achieve the goals of a program.

c. At the present time, it is likely that funds for evaluative research will be expended to evaluate projects rather than programs. However, since references in the literature of evaluative research are generally to programs, this term will be used interchangeably with project.

4. Social action or social intervention programs -- These are programs of planned social change. For example, the District of Columbia Prison College Project is an example of planned social change. This program enables approximately 150 men assigned to Lorton to enroll in accredited freshmen and sophomore Federal City College courses (Comprehensive Plan, 1974:10).

Purpose of Evaluative Research

"The purpose of evaluation research is to measure the effects of a program against the goals it set out to accomplish as a means of contributing to subsequent decision making about the program and improving future programming" (Weiss, 1972:4). This definition of purpose takes into account the measurement of effects of a program. The concern then is with research methodology to ascertain outcomes. Implicit in the comparison of effects with goals is the use of criteria to determine how well the program is doing, and, according to Weiss, the contribution to subsequent policy decisions and future program improvement signify the social purpose of evaluative research.

When an evaluation of a program is undertaken many persons will expect feedback of the evaluation findings. First, there are the top policy makers, including those in the executive and legislative branches (if the program is publicly funded) who need to know whether to continue, change, or drop the program. Directors of programs need to have information on how well the program is meeting its goals, how efficient it is, and what program alternatives are available if required. Funding agencies, in our case LEAA and OCJP&A, may have some concerns with the theory and methods of programs. For example, are work release programs instrumental in reducing crime among participants in the program? If there are indications that the effects are positive, which programs are most efficacious? Planners in the funding agencies are also interested

in feedback from evaluative research since this provides them with the necessary information for the accomplishment of planning and program analysis prior to program implementation. Results of evaluative studies need to be disseminated to interested professionals in other communities who are involved in similar intervention programs. When communication is facilitated it permits knowledge to become cumulative and prevents the same programs from being tested over and over again (Brooks, 1971:58). Finally, the American taxpayer who is the ultimate source of funding is entitled to be informed as to the value being received for the dollars spent.

Evaluative Research: A Phase In Program Development

Evaluative research can be viewed as a phase in program development. In the ideal situation, the following steps occur in the planning phase (Caro, 1971:3-4).

1. Identification of problems
2. Specification of goals
3. Analysis of the causes of problems and the deficiencies of current programs
4. Consideration of program alternatives in terms of goals, costs, and budget ceilings

Evaluative research follows program implementation and is concerned with program execution although frequently it is planned prior to implementation. The steps in an evaluative research study overlap in the research

process. Therefore, these steps or activities do not follow exactly the sequence which is outlined below:

1. Formulating the evaluation question in terms of whether the program is successful in reaching its goals
2. Formulating the program goals
3. Specifying and defining:
   a. Independent or "causal" variables which are the program inputs
   b. Intervening variables, i.e., the factors which mediate between inputs and outcomes
   c. Dependent variables which are the outcomes, i.e., the program outputs or program performance.
4. Specifying, describing, and monitoring the program
5. Designing the study.
   a. Selection of the population to be studied, e.g., program population vs control population
   b. Timing of the investigation
   c. Establishing the procedures for collection of data. The decision to examine one project or compare several projects with essentially similar goals is made at this time. Measures to be used are (1) developed at this time, (2) selected from among standardized instruments or scales, (3) borrowed from other evaluation studies of similar programs. A combination of various measures may be used.

    d. Cost - benefit analysis (if it is to be included)

6. Collection and preparation of the data for analysis

7. Analysis of the data including comparison of program effects ~~with program effects~~ with program goals

8. Interpretation of the findings and recommendations for action

9. Written report of the evaluative research study

10. Dissemination of report to:

    a. Policy makers and legislators

    b. Program directors

    c. Funding agencies

    d. Planners

    e. Professionals and organizations involved in similar action programs

    f. Public

The planning-action-evaluation cycle described above may be repeated *indefinitely until objectives are reached or until problems and objectives are redefined.* ~~indefined.~~ The most desirable method of conducting evaluative research is to build it into a new program. This is called formative evaluative research. For many reasons this is not always possible. When evaluative research appraises the results of stable and well-established programs; then it is called summative evaluation.

Program Evaluation Models

In seeking to conceptualize an approach to evaluation, the investigator makes use of a research model. Among the two most important are the goal-attainment model and the systems model. A perusal of the

literature of evaluative research indicates that both from a theoretical and methodological point of view, a number of issues have been raised relative to the interpretation and feasibility of the two models.

Of the two perspectives, goal-attainment is most favored among those who are currently engaged in evaluative research. Schulberg and his colleagues (1969) who advocate the systems model were able to include only one article based on that perspective *in the book which they edited.* With the exception of that one article, all the other evaluative research studies reported in their volume utilize the goal-attainment model. A recent special monograph of Evaluation (a new journal in the field) is devoted to describing four different methods of program evaluation in mental health settings using the goal-attainment model.

When the goal-attainment model is employed, evaluation is conceived as measurement of the degree of success or failure sustained by the program in reaching previously set goals. Therefore, the first task of the evaluator is to determine the goals of the program. This is frequently a difficult thing to do since program administrators, in answer to a question on what the program is trying to accomplish, will give fuzzy answers and sometimes will enunciate goals which are global and unrealistic; e.g., to reduce crime, to improve education, to build character, etc. When this occurs, the evaluator must guide the administrator to state the goals so that they are clear, specific, and measurable (Weiss, 1972:26). This collaborative process, then, is one of clarification of the goals so that they are well-defined, unambiguous, and can be operation-

alized for measurement using the many techniques available to the researcher. These techniques include observation, search of existing records, interviews, questionnaires, sociometric choices, laboratory experiments, physical examination, etc.

The evaluator who utilizes the goal-attainment model must be aware of some of the problems relevant to this approach and must, therefore, guard against the following:

1. Premature program conclusions -- A program which is evaluated at an early stage can be unfairly marked as a failure at a time when it is still groping for direction. Caution needs to be exercised so that a program not be prematurely discarded if it is not demonstrating scheduled achievement early in the program.

2. Unrealistic goals -- Care must be taken that goals are realistic. When goals are used as the yardstick against which program performance is evaluated, unrealistic or idealized goals may make real accomplishment appear insignificant.

3. Publicly stated goals -- Very frequently, publicly stated goals are not the real program goals. If the evaluator uses these goals, he will surely fail the program and administrators will just as surely accuse him of having used the wrong goals. So, the experienced evaluator must search for the covert goals, those which are not likely to be articulated, and use these as the standard with which to compare accomplishments. In some juvenile

correctional institutions, for example, the public goal is remedial treatment because that is what the public expects. But because of many problems, including insufficiency of resources, the real goal is mainly custodial care. If analysis of effectiveness is related to the public goal of rehabilitation, then the program is a failure, but if it is related to the covert goal of custodial care, then it is likely to be considered successful.

4. Evaluating performance in terms of the goals without determining whether the program, organizational factors, or environmental (i.e., community) factors impinge upon the outcome. If the evaluator measures only program outcomes without examining inputs into the program and what happens to these inputs (i.e., thruput), he will be unable to explain how and why the program is accomplishing its goals. In the event that it is not, he will not be able to recommend alternative ways to meet the objectives.

The systems model is an alternative approach to program evaluation. "The starting point of this approach is not the goal itself but a working model of a social unit which is capable of achieving a goal. Unlike a goal or a set of goal activities it is a model of a multifunctional unit" (Etzioni, 1960:261). In addition to achievement of goals, the other functions of "social units" (i.e., organizations within which social action programs are conducted) are: adaptation of the organization to the environment and to its own internal demands, acquisition and

maintenance of necessary resources, and effective coordination of organizational subunits. The following is a brief discussion of these functions:

1. <u>Adaptation of the organization to the environment and to its own internal demands.</u>

   a. Most systems theorists believe that organizations adapt to their environment, i.e., to other organizations, government agencies, the public, etc. The following is an example: Somewhere in the environment there are sub-publics (e.g., well-educated people, criminologists, psychiatrists, civil libertarian lawyers, universities, etc.) who believe that the most important goal of "State Prison" should be rehabilitation using various therapies. However, these groups are far less powerful than those in the community who stress that it is far more important to protect members of the community by taking strong security measures to prevent escapes from the institution. Since funds are limited and it is not the function of corrections officials to change attitudes in the community, "State Prison" adapts to the situation, i.e., the environment, by expending a greater proportion of its funds for custodial care and security rather than for rehabilitation.[3]

   b. An example of adaptation to internal demands may be seen in the recent "shakedown", i.e., a massive search for weapons at Lorton, in order to avert a walkout of guards (Washington Post, 1973: B1 and B8).

2. <u>Acquisition and maintenance of necessary resources</u> -- This function requires that effort be expended to acquire and maintain funds, personnel and other resources that insure the functioning of the organization. Thus, what this entails is "a balanced distribution of resources among the various organizational needs, not maximal satisfaction of any one activity, even of goal activities" (Etzioni, 1969:263).

3. <u>Effective coordination of organizational subunits</u> -- Just as the organization adapts to external (i.e., environmental) conditions to function so there are internal conditions which permit the organization to function. This means that the various parts of the organization must be integrated in order that the organization function effectively. If the various units are not effectively coordinated, the organization will be unable to maximize its adaptation to external and internal conditions; it may have difficulty acquiring and maintaining funds and other resources; and, finally, it will have difficulty achieving its goals.

4. <u>Goal achievement</u> -- There are several questions that need to be answered in a systems evaluation study. These are:

   a. What are the actual goals of the program?

   b. What are the public (overt) and private (covert) goals of the agency which is implementing the program?

   c. Which of these are the true goals?

d. Are these goals compatible so that program goals become

organizational goals?

Assume that these questions are satisfactorily answered. Assume also that the evaluator has immersed himself in the functions (described above) of the organization or agency which is implementing the program. He must now clarify the true goals so that they are <u>clear</u>, <u>specific</u>, and <u>measurable</u>. This is the same task which the evaluator following the goal-attainment model must accomplish. From here the "roadmaps" for pursuing a systems evaluation are very general and do not provide the evaluator with a guide to the study of the effectiveness of a social action program. Let us see what Georgopoulos and Tannenbaum(1957: 535-536), whose systems evaluation of organizational effectiveness is one of the most cited in the literature, tell us about conducting such a study. They "define organizational effectiveness as the extent to which an organization as a social system, given certain resources and means, fulfills its objectives without incapacitating its means and resources and without placing undue strain upon its members." The following general criteria are subsumed under the concept of effectiveness:

a. Organizational productivity; i.e., movement of the organization

toward its goals.

b. Organizational flexibility in the form of successful adjustment

to internal organizational changes and successful adaptation to

externally induced change.

c. Absence of intraorganizational strain, or tension, and of conflict

between organizational subgroups.

These, then, are the criteria for operationalizing organizational or program goals to make them measurable. But these criteria, despite their generality, do not provide the evaluator with a methodology for pursuing a systems analysis of social intervention programs.

The systems model in evaluative research is also suggested by Weiss and Rein (1971). They propose its use in evaluation of broad-aim programs, i.e., social action programs which are concerned primarily with impact on a situation and only secondarily with the impact on individuals. They offer a methodology for analysis of the dynamics of agency operation and program implementation but they do not know how to determine program effects on persons and institutions. The only way to develop such a methodology, they say, is to undertake evaluation research.

There are of course, other problems than those of a methodological nature in conducting evaluative research based on a systems perspective. First, it is necessary for the evaluator to have an extensive knowledge of the organization which is carrying out the program. He must also have an understanding of the optimal allocation of resources among organizations, maintenance, and goal-attainment functions (Schulberg and Baker, 1971). Second, the cost in dollars of a systems perspective evaluation is very high, and third, the time spent in conducting a systems analysis

is considerable.

Weiss takes the view that the goals-attainment model and the systems
model are complimentary.

She writes that:

> As we learn more about implementation, we can begin to
> identify vital elements in the operating systems and move
> toward description and measurement of them. In time we
> can combine the study of program process with the study
> of outcomes. In the interim, it is not unimportant to
> know how the intended beneficiaries of the program are
> faring.

Suchman (1967:61-66), whose evaluative research is based on the
at all
goal-attainment model, does not discuss the systems approach in his

book on evaluative research. However, he proposes that the _process_

of the social action program in goal-attainment should be one of five

criteria by which the success or failure of a program may be evaluated.
Criteria   *adequacy of performance*
The other four are effort, performance, and efficiency. The following

is what Suchman has to say about process:

> In the course of evaluating the success or failure of a program,
> a great deal can be learned about how and why a program works
> or does not work. Strictly speaking, this analysis of the process
> whereby a program produces the results it does, is not an inherent
> part of evaluative research. An evaluative study may limit its
> data collection and analysis simply to determining whether or
> not a program is successful according to the preceding four criteria
> without examining the why's and wherefor's of this success or failure.
> However, an analysis of process can have both administrative and
> scientific significance, particularly where the evaluation indicates
> that a program is _not_ working as expected. Locating the cause of
> the failure may result in modifying the program so that it will
> work, instead of its being discarded as a complete failure.

III. ISSUES AND PROBLEMS IN THE METHODOLOGY OF MEASUREMENT AND DESIGN OF EVALUATIVE RESEARCH

> The scientific method is not bound by either subject matter or
> objective. Hence, evaluative research has no special methodology
> of its own. As "research" it adheres to the basic logic and rules
> of scientific method as closely as possible. Its canons of "proof"
> and its laws of inference are the same as those of any research
> project. It utilizes all available techniques for the collection
> and analysis of data, and employs a wide variety of research designs.
> It may be carried out under experimental laboratory conditions or
> in the natural community. In other words, evaluative research is
> still research and it differs from nonevaluative research more in
> objective or purpose than in design or execution (Suchman, 1967:
> 81-82)."

Suchman and many other evaluation researchers make the point that

far too much evaluative research is conducted by people who have no

training in research methods. This state of affairs is unfortunate

since social programming which is designed to produce some desired

change "is the main form of scientific research for the testing of

administrative principles. Evaluative hypotheses are largely admini-

strative hypotheses dealing with the relationship between some program-

matic activity and the attainment of some desired action objective"

(Suchman 1967:133). Thus, the objective of evaluation is to increase

the effectiveness of program administration.

An evaluator's technical competence in research does not imply *that*

methodological problems will not be present in evaluation. Because of

the nature of our social problems and the complex society in which we

live, our reforms do not offer easy solutions. Consequently, social

action programs effect only small amounts of change and in order to
demonstrate such small amounts of change, measurement must attain
very high precision. Let us look at education as an illustration.
The great changes in this area have already taken place. Most of the
population is literate, the majority has at least a high school edu-
cation, and a substantial minority has attained four or more years of
college (Educational Attainment, 1970:1-3). These gains were demon-
strated with simple measures. Now changes which were once massive
occur in very small increments and consequently measurement must be
more precise. Now, let us take as an example a program to increase
the reading skills of children of migrant farmers. Augmenting the
reading skills of such a group would undoubtedly be difficult because
the population is highly mobile and for many reasons is resistant to
educational improvement. Consequently, since the degree of change, if
any, is likely to be small, researchers must develop very precise
measures of effectiveness of such a program. Such measures will, of
necessity, be very costly to develop both in dollars and in time.

In Part II, I outlined the steps that an evaluator would take in
conducting an evaluative research study. Some of these steps are
discussed in this section. Information on others can be found in social
research methods texts and in the evaluative research literature.

*demonstrates basic concepts of
sci. mths & research -*

Formulation Of The Evaluation Question and Evaluation Criteria

"The traditional formulation of the evaluation question is: To
what extent is the program succeeding in reaching its goals?" (Weiss,
1972:24). Once the research question has been stated, the next step
is to determine the program goals and state them so that they are clear,
specific, and measurable. This was discussed earlier in the section on
the goal-attainment model. In the process of clarifying the goals of a
program, criteria are identified. These criteria are used to evaluate
performance against objectives. For example, if a governmental objec-
tive is 'to reduce crime,' then it is appropriate to use crime rates
as the major criterion (but not necessarily the only criterion) for
evaluating activities aimed at this objective (Hatry, 1967:5). However,
in identifying evaluation criteria, the evaluator must take into account
the customer of evaluative research who will be making decisions based
on the evaluation. In view of this, Coleman (1971:282) writes that
"...a successful evaluation must focus on those criteria on which the
customer wishes to base his choice. This is perhaps the most crucial
element in the design of evaluation research for unless the appropriate
criteria are used, then the results are irrelevant to the choice that
must be made .... One reason that the step of determining the criteria
to be used in evaluation is so crucial is that often the customer him-
self is not fully aware of the criteria he wants to use."

Ideally criteria for use in evaluative research should have the
following properties (Hatry, 1967:6):

1. Each criterion should be relevant and important to the specific problem for which it is to be used.

2. Together the criteria used for a specific problem should consider all major effects relative to the objectives. Enough criteria should be evaluated to cover all major effects.

3. Each of the criteria ideally should be capable of meaningful quantification. However, because of limitations of many current information systems, it may on occasion be necessary to utilize qualitative criteria.

Program Goals

Hatry points out that the process of selecting criteria will often suggest the need for revising goals. It is at this particular juncture, then that goals of the program will need to be reexamined to determine whether they are indeed appropriate. "To what avail is the result of program evaluation if wrong targets had been chosen at the outset of the service?" (Greenberg, 1971:155).

Many researchers distinguish among immediate, intermediate, and ultimate objectives of programs. This distinction is important because government agencies and foundations frequently require that demonstration projects be evaluated. Since such projects usually do not have the form

that in time they will develop as they are perfected, early evaluation may tell little about their potential effectiveness. Therefore, evaluators must exercise caution in presenting results at a very early stage in order to prevent promising programs from being scrapped. According to Greenberg (1971:158-163) some immediate goals of a program can be measured within six to twelve months of commencement. However, those goals will not _these_ likely represent input into the system as can be seen from the example of immediate goals of a heroin enforcement program given below. Inter-mediate goals concentrate on early benefits of a program and in the case given _below_ this would refer to the large numbers of addicts in treatment and the decreasing potency and availability of heroin. Greenberg believes that it takes between three to five years for intermediate effects to appear. On the other hand, the efficacy of long range goals may not be discernible before a period of ten years has passed.

An example of the possible outcomes of a heroin enforcement program which includes enforcement, research, and rehabilitation components might be as follows:

1. Immediate goals -- large numbers of arrests of heroin pushers and dealers; pharmacology research program set up to develop non-habituating drugs to help addicts break heroin habit; urine testing program of all arrestees; addicts urged to enter existing drug rehabilitation centers; implementation of method for surveillance of all addicts assigned by the Criminal Justice

System to drug rehabilitation centers; development of method for screening out non-addicted occasional heroin users in methadone maintenance programs; counseling, educational, job-training, and job placement, programs made available to clients of drug rehabilitation centers.

2. Intermediate goals -- arrests of higher level heroin dealers; significant decline of amount of heroin available on the street; decline in quality, i.e., potency of heroin; large numbers of addicts in treatment centers; promising non-addictive drugs that would inhibit desire for heroin now in the experimental stage; substantial percentage of program participants working or enrolled in job training courses.

3. Long-range goals -- significant decline in certain types of crime, e.g., shoplifting and burglary (other factors may also be responsible for this development); rehabilitation of former addicts; significant decline in estimated numbers of addicts; several non-habituating drugs developed which were tested on humans and were successful in breaking the heroin habit -- these drugs now being used in drug rehabilitation; former addicts who used methadone to break heroin habit, successfully completing program to terminate use of methadone; substantial percentage of program participants self-supporting.

Research Design Of The Evaluative Study

The design, i.e., the plan of study, of an evaluation depends on several factors. Among these are available funds, point in time when evaluation must be completed, type of program being evaluated, and degree of knowledge currently available about the program. Taking these factors into account, the following research designs are among the options available to the researcher:

1. The experimental design -- This is the classic design for evaluative research. Individuals are randomly chosen from the total population in a social action program, to be in either an experimental or control group. The experimental group receives the program, the control group does not. In a program to increase reading ability, for example, measurements are taken of the experimental classrooms and control classrooms before the program starts and after the program ends. Comparisons among the groups are made to determine whether changes have occurred. If improvement in reading ability among children in the experimental classrooms is significantly greater than among children in the control classrooms, the program can be considered a success. In this case, children in the control classrooms are receiving the regular reading program since it is not possible to discontinue reading lessons. Unfortunately, this design is not always feasible in action settings because there may be too few people to act as controls in a program which serves most of those who are eligible. Another problem in the use of the

experimental design is that program personnel may feel
that it is unethical to have people in the program and
then deny them service by placing them in a control group.
One way to get around this problem is to substitute for
the control group, a group receiving a comparison social
program. Thus, two experimental groups are used in order
to learn which program is superior. Another solution,
where a program is still experimental, is to provide all
communities in a program with a variety of treatments at
different times and compare them. As a research design,
the experimental method is very rigorous and is generally
not applicable to evaluation of programs about which
little knowledge is available. According to Schulberg
and his associates (1969:12), "Less rigorous designs
can be appropriately employed when ambiguity is great,
while more rigorous designs should be utilized as
knowledge about relevant variables increases."

2. Quasi-Experimental Design -- When program conditions
rule out the experimental design, the quasi-experimental
design is frequently utilized by evaluators. "The basic
criterion for how satisfactory they are is the extent to
which they protect against the effects of extraneous
variables on the outcome measures. The best designs are
those that control relevant outside effects and lead to

valid inferences about the effects of the program....
Quasi-experiments have the advantage of being practical
when conditions prevent true experimentation" (Weiss,
1971: 67-68). Among the many quasi-experimental designs
are the following:

   a. Time-series design in which a series of measure-
ments are taken at periodic intervals before the
program begins and continues on to a series of
measurements following completion of the program.
These measurements provide the researcher the
opportunity to determine whether later measurements
indicate a marked change over earlier patterns prior
to program implementation. An example of this design
is the following: Measurements of functionally
illiterate prison inmates who are to be in a program
to improve reading and writing skills are taken
several times prior to implementation of the program,
several times during the program, and several times
following completion of the program. Later measure-
ments can be compared with earlier measurements to
determine the degree of change in literacy; Also,
the last measurement can be compared with the one
taken immediately following the end of the program
to ascertain whether the new skills are retained

by inmates completing the program.

b. Multiple-series design which incorporates the features of the time-series design but takes measurements of a control group, i.e., a similar group or institution which does not receive the treatment.

An interesting variation of the quasi-experimental design was carried out by Wright and Hyman (1964) in an evaluation of the Encampment for Citizenship, a summer institute for training young people in responsible, democratic citizenship. This was done by incorporating the design into two waves of self-administered questionnaires, varying wordings of several questions in different sets of questionnaires, and controlling the assignment of these different forms to various subgroups in the program.[4]

There are times when even the quasi-experimental design cannot be used. This may result from limited time allowed for an evaluation, lack of availability of a non-random control group, or lack of funds. The problem with non-experimental designs is that they cannot pinpoint whether observed changes are due to the program or to alternative causes. Three of the most common non-experimental designs are the following:

1. Before-and-after study of a single program in which measurements of the participants are made to determine how well they have responded to the program.

2. After-only study of program participants. This is an ex-post-facto design in which retrospective reports on prior status of participants are collected and in turn are compared with the "after" status of participants.

3. After-only study of participants and non-random "controls." The latter are a comparison group who are selected because of their similarity to program participants.

Measurement: Indicators, Reliability, and Validity

According to Suchman (1967: 116) "public service and social action research are deficient in their concern with the reliability and validity of...evaluative instruments." Apparently, the criteria in evaluative research (discussed earlier in this section) create measurement problems similar to those of operational indices for non-evaluative research. These problems are connected with choosing indicators to be used in measurement which will serve as a link between the criteria and program objectives. Lazarsfeld and Rosenberg (1955: 15) write about this problem in their book, The Language of Social Research. They deal with two basic questions:

1. How does one "think up" indicators for the criteria (variables) being studied?

2. How does one select from all possible indicators those which are "part of" the criteria and, therefore, will be measures of the criteria.

Indicators of the criteria will depend on how the criteria are defined, how much time is spent on clarification of the criteria definitions, and the precision of measurement of the criteria which is desired. Let us go back to Hatry's example (1967:5) that I gave earlier in which a governmental objective "to reduce crime" was identified and "crime rates" are the major criterion.

How should crime rates be defined? The 1972 issue of Uniform Crime Reports gives the following definition: "Crime rates relate the incidence of crime to population. A crime rate should be considered a victim risk rate in that it demonstrates the risk of becoming a victim of crime." Using this definition, which indicators shall we select? We can use an overall crime rate which tells us that in 1972 there were 2,829.5 crimes per 100,000 residents in the United States. We can also provide for 1972 a crime rate for each of the crimes included in the Index of Crime. If the definition of crime rate is

further clarified, indicators of crime rate, e.g., region, urban-suburban-rural, age, sex, race, socio-economic status, etc. might also be included. These variables will indicate that the risk of becoming a victim of crime will vary for different groups of citizens. A very important variable which is not included in official crime rates is the rate of unreported crime. How does one estimate this? LEAA is presently collecting *survey* victimization data and probably in time this new source of data will be utilized to obtain an estimate of crime which is closer to the true crime rate.

At this point it becomes necessary to further clarify the goal "to reduce crime." Does this goal mean to reduce crime by a small amount or a great deal? Are we concerned with reducing the incidence of violent crimes? By what rate? Shall be compare crime rates for 1972 with crime rates for 1971 and show that the overall percent change in rate was a decrease of 2.7 percent? Does the overall rate accurately reflect individual crime rates? If we are concerned with the crimes of murder, forcible rape, and aggravated assault we discover that these crimes have increased since 1971 while robbery, burglary, larceny $50.00 and over, and auto theft have shown decreased rates over 1971. Does the governmental objective "to reduce crime" mean that the baseline year should be 1960? In

1972 the increase of the national crime rate over 1960 was
151.2 percent. During this period the murder rate increased
by 78 percent, forcible rape by 134.7 percent, and robbery
by 200.3 percent. Other crime index rates also showed
comparable increases. If we decide that reducing crime
to 1960 rates is not a reasonable goal in terms of
present social conditions, then the use of 1967 as a baseline
year may be more feasible. Between 1967 and 1972 the
national crime rate increased by 46.9 percent, murder by
45.9 percent, rape by 61.6 percent and aggravated assault
by 34.5 percent. The remaining crime rates in the 1972
index of crime also showed comparable increases over the
1967 rates.

Whether the indicators incorporated into the
measuring instruments (be they agency or primary data)
demonstrate the connection between criteria and program
goals will depend upon the precautions taken by the
evaluative researcher to insure reliability and validity
of these measures.

Reliability in evaluative research is far more
problematic than in other types of social research. By
definition, reliability refers to the degree to which
consistent results can be obtained on repeated applica-
tions of a measuring instrument, i.e., that the same

results will be obtained each time the instrument is
used. Thus, the focus of reliability is on the way
measurement is accomplished rather than on what is
measured. One way to insure reliability of the measuring
instrument is to make it as precise as possible. To
further insure reliability of a measure when ratings or
codings are involved, steps must be taken to determine the
consistency between raters and also by the same rater at
different times. These steps are very important in
evaluative research because the evaluator is attempting
to determine whether actual change has occurred. Poor
reliability, i.e., variation in measuring instruments,
coding, etc., may either obscure real change or may indicate
change when in fact there is none.

An example of reliability of a measurement might
occur in a drug treatment center where urine specimens are
collected on a regular basis from persons in treatment.
The urine specimens are sent to a laboratory to undergo
chemical analyses by thin layer chromatography to provide
positive identification of persons who have used drugs
quite recently. How reliable are the laboratory
measurements? One monitoring procedure to check on
reliability is to submit split or replicate samples, i.e.,
to divide actual samples and submit each subpart to be

analyzed as a separate sample. This procedure which was
carried out in a recent study on drug usage of arrestees
provided a reliability check on the technique of chemical
analysis according to the researchers who stated, "This
monitoring effort provided information concerning the
consistency of the laboratory procedures, that is, agree-
ment between different samples of the same specimen was
accepted as an indication of a consistent laboratory
procedure. However," the authors continue, "this provided
little information on the accuracy of the reports"
(Eckerman, et. al., ~~B6-137~~ *1971:86*).

When accuracy is mentioned, attention shifts from
reliability of the data to validity of the data. The
concern is now with what is measured rather than how it
is measured. Using the same study discussed above as an
example, let us see how the researchers validated their
findings. A research chemist on the project prepared
standard solutions of five drug materials which the
chemical analyses in the laboratory were supposed to
detect. The solutions were added to stock urine in
sufficient quantities to be detectable. These samples
were included with the urine specimens collected from
arrestees in the study and sent to the laboratory. "The
purpose of the monitoring program was to provide an

indication of the ability of the laboratory to detect the
presence of the drug substances when they are known to be
present. This information would then be used to assess the
validity of the urine analysis reports for the total
specimens obtained from men in the study" (Eckerman, et. al.,
1971: B-9).

There were six validity checks made in succession, one
for each city in which the study was conducted. In the
first test the "laboratory correctly identified 93.1
percent of the standard samples containing morphine (the
drug used to determine detection of heroin in actual samples)."
As a result of this finding, procedures for identifying
morphone (heroin) were improved resulting in almost 100
percent detection in the next five sites.

There was still another validity test used in this
study. Each respondent was asked a series of questions on
drug use. These data were compared with the results of
the chemical analyses. The results indicate a strong
correspondence between the laboratory findings and the
interview data.

Now I would like to go back to the example of the
drug treatment center. Let us assume that chemical analyses
of urine specimens of patients are both reliable and valid
and patients at the center are truthfully reporting that

they are no longer using heroin. Are the patients success-
ful because they are now heroin free? Is being heroin free
a valid indicator of success? Should being crime free also
be included as an indicator of success? The individual
who is being rated by treatment personnel may be on welfare.
Is he successful if he is heroin and crime free but cannot
earn a living? Suppose the patient learns an occupational
skill but does not obtain a job. Is this because he really
does not want to work or is it due to the hesitancy of
employers to hire recent drug addicts? How do you measure
success now? Let us assume that the patient does obtain a
job for which he is trained. Does this mean that he can
provide for himself and his family? How long must he
retain the job to prove himself? If he uses methadone
which he receives from the treatment clinic, is he success-
ful if he no longer uses heroin, is gainfully employed,
supports himself and his dependents, and leads a life
which is consonant with the mores of his society?

How valid are the above indicators of success?
Herzog (1959: 45-46) indicates that "in assessing psycho-
social change, the question of validity rests ultimately
on opinion." As can be seen from the examples given above,
tests of validity depend upon the type of scientific
measure to be validated. Given a laboratory situaiton and

adequate controls, validity can be based on determined fact.
Yet even in the laboratory, validity depends on how adequate
the test is and the efforts taken to prevent testing errors.
Sometimes tests such as the chemical analyses of urine
specimens are available to determine whether respondents
are answering questions on drug use truthfully. In the
case of psycho-social change, the ultimate criterion of
validity is informed opinion. These are some of the ways
in which tests of validity of measuring instruments are
used to determine whether they do in fact measure what they
purport to measure.[5]

The type of measuring instrument that is used in an
evaluation study depends on the purpose of the social
action program. According to Weiss (1972: 39) such
instruments can be concerned with attitudes, values,
knowledge, behavior, budgetary allocations, agency service
patterns, productivity, as well as many other groups of
variables.

Agency Data Versus Data Collection

Shortcomings of agency data usually present problems
to the evaluator who seeks valid, reliable, and sensitive
measures. Lerman (1971: 221-232), for example, discusses
problems of institutional records of delinquents that

reflect organizational interests as much as they do behavior
of inmates which they purport to measure. In some private
institutions supported by public funds, delinquents who are
judged to be untreatable are transferred to public
institutions. However, the social bookkeeping method
which these institutions use, treat these internal failures
as nonexistent by not showing them when the final figures
are tabulated. Only those residents who complete treatment
and are released by the institution are included. Success
of the institution is judged by following up only those who
complete treatment while the failures are regarded as if
they_ never been part of the institutional population.

Because of these and similar problems researchers must
often collect their own data. While this procedure has
obvious advantages, it may result in a new set of problems.
These are:

1. Data collection may add substantially to the cost
of evaluation.

2. Administrators and practioners may feel that data
collection interferes with their work.

3. Evaluators become concerned that data collection
may artificially enhance client awareness of the program,
thereby, changing its effectiveness. There are, of course,
ways of dealing with such problems as, for example, using
unobtrusive measures to collect data (Webb, et. al., 1966).
There are many types of data available to the evaluative
researcher depending upon the type of program, available
resources, and ability of the investigator to utilize various
research techniques. These are interviews; questionnaires;
observation; ratings (by peers, staff, experts);
psychometric tests of attitudes, values, personality,
preferences, norms and beliefs; institutional records;
government statistics; tests of information; skills;
application of knowledge; projective tests; situational
tests presenting the respondent with simulated situations,
diary records; physical evidence; clinical examinations;
financial records; and documents (e.g., minutes of Board
meetings, newspaper accounts of policy actions, transcripts
of trials) (Weiss, 1972: 53).

Input, Thruput, and Output

Although most evaluative researchers use the goal-
attainment model many are concerned with tracking the
process of the program from input to outcome. Why should

the evaluator who is using the goal-attainment model be
concerned with tracking the program? After all, the
evaluator's job is to determine whether the program is
achieving its goals? Why bother with what goes into the
program? Can the evaluator assume that the statement of
the program plan as it appears on paper has actually been
implemented by the staff? If he does, there is always the
possibility that he may be observing the effects of a non-
existent stimulus. For example, suppose thousands of
posters with motivational appeals to heroin addicts to
register for treatment at drug rehabilitation centers were
slated to be widely d'stributed by volunteers to the
population of several high drug use communities. However,
the volunteer project was not organized and the posters were
collecting dust in a storeroom. An evaluator, unaware that
the treatment had not been applied might give the program
failing marks and then blame it all on the inadequacy of
communications and motivational theories.

Besides being certain that the program is really
happening, the evaluator has to find out what is actually
taking place. If a service is offering job training to
clients recently released from prison, then the evaluator
will need to monitor the program to determine what kinds of
job training are being given, what qualifications teachers

possess, where the training is taking place, the adequacy
of training materials, length of time of training, etc.

When the evaluative researcher investigates what goes
into a program he is looking at such inputs as services,
staffing, persons served, staff hours worked, management,
funds (more will be said about this in the section on
cost-benefit analysis), etc. If the participants are
viewed as inputs, then their characteristics can also be
classified as inputs. These may include age, sex, race,
socio-economic status, length of residence in community,
attitudes toward the program, etc. These inputs can be
grouped as (1) the administrative pattern, (2) the service
pattern, and (3) personal characteristics.

An important area for investigation is what I call
thruput. This is the region between inputs and effects
(outputs). Here is an example that might occur. Inmates
in a prison system show considerable discontent by complain-
ing about the poor quality of food which they receive. An
evaluator discovers that prison officials are expending
large sums of money for food of adequate quality. What,
then occurs between the time the order for food is sent to
the food wholesalers and the time the prepared food is

served to the inmates? In the course of the investigation,
the evaluator discovers that (1) food of a lower quality
than is ordered and paid for is received and accepted at
the prison; (2) storage facilities are poor resulting in
food spoilage; (3) there is no dietitician on the staff to
plan meals and supervise cooking; and (4) food handlers and
cooks are not trained in their duties. In this example
there is a great discrepancy between inputs as disbursed
and inputs as services received. It is, therefore, very
important in evaluation to examine what happens between
the time of inputs as disbursed and inputs as received.
Coleman(1971: 284) states that "it may well be that the
principal, or at least a major, explanatory variable in
the effectiveness or ineffectiveness of a given program
is the loss of input between its disbursement by
authorities and its reception by (those) it is intended
to effect."

Output has to be considered not only in terms of goal
fulfillment which I discussed earlier but also in terms of
unanticipated effects which may or may not be favorable.
An example of unanticipated program output is the following:
A program in which juvenile delinquents receive service,
benefits the community as well as the delinquents by
reducing vandalism and other juvenile crimes. This is a
beneficial unanticipated side effect which was not targeted.

However, in its initial phase, the program creates
considerable ill feeling, divisiveness, and actions to
scrap the program because administrators attempt to set up
a juvenile center in an upper income residential area of
the community. This is an unintended and unanticipated
consequence of the program.

Cost-Benefit Analysis

Having determined whether the goals of a program
have been achieved, the evaluator now proceeds to an
analysis of the benefits versus the costs of the program.
This is referred to as a cost-benefit analysis and can be
shown as an Efficiency Index (Greenberg, 1971: 161) as
follows:

$$\text{Efficiency} = \frac{\text{Output (in terms of goal fulfillment)}}{\text{Input (in terms of dollars, services and/or Personnel time)}}$$

The most efficient program, then, can be defined as
that which yields the greatest per unit change not the one
that can be run at the least cost-per recipient (Freeman
and Sherwood, 1971:272). However, cost should receive
particular consideration in circumstances where alternative
programs yield similar effects. There are times, also,
when programs which are not effective must be continued
because the ideological climate in the community supports

them. The evaluator, then, has the responsibility of comparing costs of various programs and recommending the continuance of only those which are most economical.

The efficiency index or ratio shown above seems like a neat measure. But how do you quantify intangible effects or benefits of a program? What is the monetary value of feelings of high self-esteem among a group of ex-convicts who have learned a trade? How do you quantify the decreased anxiety of parents whose children are no longer involved in vandalism? "Cost-benefit analysts have to cope with this problem by acute selection of indicators of benefit and by qualitative reporting of factors not susceptible to monetary expression" (Weiss, 1972:86).

For some types of programs in which benefits are quantifiable cost-benefit analysis is a logical extension of the evaluative research effort. And it is particularly useful to those in policy making positions who find that cost-benefit analysis rationalizes, clarified, and simplifies evaluative research reports.

IV. PROBLEMS IN EVALUATING INTERVENTION PROGRAMS IN SOCIAL ACTION AGENCIES

Historically, decision makers have not drawn upon evaluative research as input into development of policy for social action programming. Rather, social intervention programs have resulted from recognition of social needs and pressures for social change. Intervention programs are rarely evaluated unless the funding agency stipulates an evaluation requirement. In the case of Federal agencies which disburse vast sums of money in the form of grants-in-aid, Congress has required that evaluation of social programs be conducted.

Problems In Administration of Evaluative Research

The literature of evaluation is replete with problems in administration of evaluative research in social action agencies. Frequently there appears to be a tug of war between evaluators and social action agency representatives, i.e., administrators and practitioners.

Among the situations which create strain between researchers and agency personnel are the following:

1. The evaluator enters the organization to examine the work of the practitioner. The latter may feel threatened by the evaluator and uncertain about the undertaking of research in an area in which he sees

himself, rather than the evaluator as the expert. Both administrators and practitioners are concerned that evaluators may point up inefficiencies in the operation.

2. The role of the evaluator is to be innovative and he may recommend new ways of treating clients. The practitioner, on the other hand, see this as a technique to undermine his established authority. He wants peace and quiet and feels that he cannot carry out the program in an atmosphere of turbulence which he believes the evaluator is creating.

3. The practitioner is inclined to think that the probing of the evaluator imposes an additional burden on his already overfull work schedule since he is now required to make more complete and more frequent recording of information to be used as data for the evaluator.

4. The agency administrator sometimes finds himself in a difficult position. He may be interested in improving program output and would like to have the assistance of independent evaluators because he may recognize that his practitioner staff will not provide him with all the information he requires. On the other hand, once evaluation is in progress, findings may uncover organizational problems which contribute to program

inefficiency. Administrators may find this embarrassing and relations with evaluative research personnel may become strained.

5. There are considerable differences in the organization of time between evaluator and practitioner. The latter allots his time to clients while the researcher who does not have clients organizes his time in terms of designing research, collecting, recording, and analyzing data, and writing a report or monograph. The practitioner may regard this use of time as "living a parasitical life free from schedules and responsibilities" (Rodman and Kolodney, 1971:122).

6. The evaluator expects to get primary credit for the report since publication represents the culmination of his work. From his point of view, the practitioner feels that he should receive a great deal of the credit because he has given much time to providing the evaluator with data.

7. Tension is frequently created between evaluative researchers on the one hand and administrators on the other hand because the former "insist that they hold ultimate responsibility for research design and execution. Administrative interference with what social scientists consider to be critical issues in the design and execution of research is seriously resented" (Caro, 1971b:10).

Overcoming Problems In Administration Of Evaluative Research

Many efforts are made to overcome the administrative problems of conducting evaluation in social action agencies. Mann and Likert (1971:143-151) suggest a specific process for involving the total structure from top management down to the practitioner in the analysis of the data. First, they say, discuss the data with top agency administrators and ask them to (1) help interpret the data, and (2) plan a program to implement the findings. Then hold a series of meetings with department heads to convey to them the findings affecting their departments. These executives are also to be asked to "help interpret the data and to decide what further analyses of the data should be made to help them in formulating plans for constructive administrative actions." The meetings with department heads are to be conducted by top agency administrators who have previously attended planning sessions with evaluators.

This process will be repeated and will filter down to division heads, then to first-line supervisors and in some instances to practitioners in the work group.

This process of involvement leads to the following positive results according to Mann and Likert:

1. A high degree of participation and personal involvement in acceptance and utilization of evaluation findings.

2. Attitude changes resulting from interactions among members participating in group discussions of the findings and group decisions concerning future actions. These problems are brought into the open and opportunities are provided for their resolution.

3. Changes are more likely to occur when employees participate in self-analysis than if analysis is made by an outsider.

Many authorities in the field of evaluative research believe that one way to reduce strain is for the researcher to avoid becoming involved in open or hidden power groups within the organization. Furthermore, the researcher must under no condition accept agency management diagnoses of organizational problems as a starting point for research. Thus, by stressing his independent role in the organization, the evaluator can maintain research objectivity and neutrality.

There has been much discussion in the literature on the comparative advantages of "inside" versus "outside" evaluators. The inside evaluator is a staff member of the organization in which programs are being evaluated while the outside evaluator is a consultant who may be a member of the academic community or of a research organization. It is generally felt that when evaluation is conducted at the request of a funding agency, it is desirable that the research be undertaken by an outside evaluator since self-appraisal by inside evaluators may

be viewed as lacking in objectivity. A report on Improving Federal
Program Performance (1971:62) notes that "The use of outside groups
is particularly important in supplementing state and local program
evaluation capabilities." Outside evaluative researchers can be drawn
from quasi-governmental organizations such as the National Academy
of Sciences, non-profit institutes, and analytic research organizations.
However, if the social action agency is large and top administration
is interested in and supportive of independent evaluation, then it
may be possible to conduct objective evaluation within the agency.

If the evaluative researcher is to carry out his mandate he must
have the firm support of the administration of the institution whose
program he is examining and he must be able in an emergency to "appeal
to authority to maintain the operation that has been set in motion"
(Mann, 1971:181). In this respect it is of the utmost importance that
administrators and staff members be made aware initially of the
implications of program evaluation. In the first instance this will
be the task of the funding agency when the decision to evaluate is
made and later by the evaluator when he is selected.

To ease the problems besetting both evaluators and administrators,
the former should, at an early stage of the evaluation, make clear some
of the limitations of the contribution of evaluative research, explain
that evaluation cannot solve fundamental value issues, nor can it resolve

by itself deep-seated conflicts between administrators and staff or
between agency and clients.

Problems In Utilization Of Results Of Evaluation

"The basic rationale for evaluation is that it provides information
for action. Its primary justification is that it contributes to the
rationalization of decision making [and] unless it gains serious hearing
when program decisions are made, it fails in its major purpose"
(Weiss, 1971:136). The literature yields little to indicate that
evaluation results are actually utilized to any great extent. Indeed
many agencies either ignore or explain away evaluation findings and are
not above casting aspersions on the researcher's understanding of the
situation and the state of the art.

Writers in the field of evaluative research discuss many reasons for
the failure to utilize evaluation results. Among these are:

1. Basic lack of understanding by administrators of the role
   of the evaluator. Administrators generally tend to place more
   credence in judgments of their department heads than in the
   scientific results of evaluators.

2. Lack of authority of the evaluator within the organization
   to conduct the evaluation. As a result, researchers find that
   they are hampered in their attempts to administer question-
   naires, to interview personnel, and to obtain other necessary

data. This problem may be overcome if the evaluator obtains the written support of the agency head (for obtaining data and accomplishing the evaluation) at the very beginning and maintains an ongoing relationship with him.

3. Where evaluation is required by law administrators may resist implementing results because of lack of understanding of the relevance of evaluation to program improvement.

4. Disagreements between evaluators and administrators regarding:

   a. Evaluative criteria.

   b. Explication of goals and strategies of the organization by the evaluator versus administrators who find it difficult or undesirable to commit themselves.

   c. Emphasis on mid-range and long-term problem solving by evaluators versus emphasis on solving immediate problems (i.e., putting out brushfires) by administrators.

   d. Commitment to scientific decision-making by researchers versus confidence in conventional wisdom by administrators.

   e. Evaluators emphasis on program effectiveness which disposes them to encourage change versus administrators emphasis on avoiding turbulence which disposes them to encourage

maintenance of the status quo.

   f. Where a program is foundering, researchers may question theoretical premises of programs, see a lack of understanding of the basic problems by administrators, and look for structural factors within the organization in a search for explanation of problems in contrast to administrators who are likely to accept the validity of the program, attribute failure of programs to inadequate application of approaches, and explain organizational behavior in terms of individual idiosyncrasies and personality problems.

   g. Underestimation of political constraints, budgetary problems, and limitations of personnel and facilities by evaluators who may be unfamiliar with administrative realities of ongoing programs.

5. Limitations of the research itself. For example, evaluative research at times may not produce results soon enough to influence short-term policy decisions.

Overcoming Problems In Utilization Of Evaluative Research

Weiss (1971:141-142) makes several suggestions in her paper to increase the utilization of evaluative research findings and help overcome some of its serious problems. These are as follows:

1. Tracking the progress of the program input through the agency makes it possible for the evaluator to inform the administrator where problems exist.

2. Analysis of components of the program and of alternative approaches allows the evaluator to recommend to the funding agency and to administrators alternative ways of programming rather than finding an entire program lacking in merit and with no indication of what course of action to take.

3. Early identification of potential users of evaluative research findings and selection of issues relevant to them.

4. Involvement of administrators and practitioners in the evaluation process.

5. Completion of evaluation on time with early release of findings.

6. Presentation and dissemination of findings using effective techniques of communication. In this respect, Mann and Likert (1971:151) recommend that the findings be formulated in simple, non-technical language and in graphical form. Going beyond communication of findings, Suchman (1967:164) makes the point that it is necessary for the researcher not only to make recommendations but to translate these recommendations into the actual organizational or procedural changes that might be

developed to implement the recommendations. "Too often, evaluation studies offer only broad generalizations about why programs are not succeeding, without attempting either to make these reasons more specific or to suggest what might be done about them."

There is another step which goes beyond use of effective techniques of communication and recommendations for social programming changes that must be taken in order to increase utilization of evaluative research. This involves the development of mechanisms for the feedback of evaluative research findings to (1) officials responsible for program planning and program modification in funding agencies and (2) administrators in funded organizations who are responsible for carrying out agency programs. This is relevant to my earlier discussion of the process of social programming in which I indicated that the feedback function is part of evaluation in the planning-action-evaluation cycle.

At the present time the prestige standing of evaluative research is low. The reasons for this state of affairs are many and varied but one of the most important is that this type of research has little impact on social action programs and is diffused to an extremely limited audience. As Rossi (1971:98) sees it, evaluation must be "accorded its proper place as playing a major role in policy formation and change. Policy is formed without considering what kinds of evaluation research would be needed to sustain the worth of a program and even more important, what are

reasonable alternatives when evaluation indicates that a program has failed. Without such a two-pronged commitment to evaluation, research tends to be wittingly or otherwise designed to produce irrelevant results shoddily conceived, poorly carried out, and easy to disregard." It would be well, then, to accord more weight to evaluative research and at the same time to select evaluators who have strong professional and organizational credentials.

V. CONCLUSION

A recent survey (Kimberling and Fryback, 1973) of the 54 LEAA funded State Planning Agencies indicates two major obstacles to the development of criminal justice evaluation systems. These are:

1. Widespread confusion over the nature and meaning of evaluation. Many SPA's consider that fiscal auditing and/or project monitoring are adequate measures to fulfill the evaluation requirement. Other SPA's regard reports of program directors and program personnel provided during site visits as fulfilling the evaluation requirement. No effort was usually made to relate these subjective opinions to program effects.

2. "Limitations on manpower and funding for the design of systems of evaluation that will render evaluative results comprehensible and ensure their timely input into the planning cycle."

As of the period between May through October 1972, according to the survey, the District of Columbia State Planning Agency did not have an evaluation plan in operation, existence, or development; we did not utilize any consultants at that time to help us in the development of an evaluation system; and our evaluation activity consisted only of project monitoring with some field team site visits planned.

Kimberling and Fryback (1973:159) believe that the development of criminal justice evaluation systems would constitute a major contribution to the national and international law enforcement community. They state that "increased LEAA requirements and funding for the development of criminal justice evaluation systems offer a significant opportunity for the design and refinement of evaluation methodologies and procedures that may prove transferrable to other large scale social programs." In the meantime, it may be useful for the program evaluation staff in OCJPA to examine evaluative research reports of grants-in-aid programs funded by the Departments of Labor, Health, Education, and Welfare, and the Office of Economic Opportunity to determine whether these studies can contribute to the developmental period of evaluation in the area of criminal justice programs.

FOOTNOTES

[1] This type of analysis was formerly known as Planning - Programming - Budgeting System (PPBS) and was introduced into the Department of Defense by Secretary McNamara in 1961. PPBS is presently in operation in some state, county, and city governments.

[2] A more sophisticated definition of program which fits into a systems model is the one given by Schulberg, Sheldon, and Baker (1969a:5) in their "Introduction" to a reader in program evaluation which they edited. They write, "In the field of organizational study, programs generally are defined as a set of activities occurring within a social enterprise which have specific inputs of resources and conditions, certain ways of organizing and processing these resources and conditions, techniques for establishing relations among them, and certain outputs which can be evaluated against given standards. Additionally, aspects of the organization's patterned activities occur not only within its own structure but also in relation to other organizations as well."

[3] In recent years considerable differences have been noted among systems theorists. Much of this has probably resulted from attacks on the theory by those who do not accept it. The latter take the position that social systems (i.e. organizations, communities, societies, etc.) do not necessarily adapt to

social change. The example of adaptation which I gave was that "State Prison" was really accommodating to the status quo in the community by placing custodial care and security first. Suppose the pressure was toward an effective rehabilitation program entailing many changes and suppose further that programs were being instituted to effectuate such changes. Could the institution accommodate to such radical social change? The social scientists who attack systems theory state that social systems do not necessarily react to social change by accommodating to it. Accommodation or adaptation implies that social systems maintain a state of dynamic equilibrium. The systems approach of Wilbert E. Moore (1963), an industrial sociologist, substitutes the postulate of a tension management system for the postulate of a system in equilibrium, thus taking social change into account. Using this perspective, the researcher can identify the tensions (inconsistencies or strains) and predict that these are the probably sites of change. This type of model differs from the equilibrium model as follows:

1. If the strains are deepseated, change will not restore the equilibrium.

2. The consequences of change will be tension producing as well as tension reducing.

3. Management of tension may not be successful. The
   system may not persist or make a transition from
   one system to another.

[4] A more complete treatment on the subject of experimental and
quasi-experimental designs can be found in Campbell and Stanley
(1966) and Campbell (1971).

[5] For the reader who is interested in further discussion of the
subject of validity, see Suchman (1967:120-126) and Herzog
(1959:41-50).

[6] Cost-benefit analysis described in this section is applied
retrospectively to calculate the return on investments in on-
going and past programs. This contrasts with cost-benefit
analysis which is conducted in a prospective framework in which
the emphasis is on planning to determine the probable costs and
benefits of alternative strategies to achieve a given end, e.g.,
to reduce crime.

VII.   REFERENCES

] P., "The Community Action Program As A Setting For
.arch," in Francis G. Caro (ed.), Readings In Evalua-
:h. New York: Russell Sage Foundation, 1971.

.ld, "Reforms As Experiments" in Francis G. Caro (ed.)
Evaluation Research. New York: Russell Sage Foundation, 1971.

.ld T. and Stanley, Julian C., Experimental And Quasi-
. Design For Research. Chicago: Rand McNally, 1966.

. G. (ed.), Readings In Evaluation Research. New York:
: Foundation, 1971a.

. G., "Evaluation Research: An Overview" in Francis G. Caro
.ngs In Evaluation Research. New York: Russell Sage Foun-
.:.

. S., "Evaluating Educational Programs," in Francis
.), Readings In Evaluation Research. New York: Russell
.ion, 1971.

Plan For Law Enforcement And Criminal Justice, Adult
.. District of Columbia, Criminal Justice Coordinating
:, 1973.

.iam C., et.al., Drug Usage And Arrest Charges. Final
. Contract No. J-70-35, Washington, D.C.: Bureau of
..d Dangerous Drugs, U.S. Department of Justice, December,

.tainment, 1970 Census Of Population, PC(2)-5B, Bureau
..s. Washington, D.C.: U.S. Government Printing Office,

.i, "Two Approaches To Organizational Analysis: A
.. A Suggestion," Administrative Science Quarterly,
2, Sept. 1960.

.d E. and Sherwood, Clarence C., "Research In Large Scale
: Programs," in Francis G. Caro (ed.) Readings In
.escarch. New York: Russell Sage Foundation, 1971.

.. "Evaluation Of Social Programs," in Francis G. Caro
.ngs In Evaluation Research. New York: Russell Sage
.971.