

EVALUATION RESEARCH:  
A TOOL FOR THE APPRAISAL OF SOCIAL PROGRAMS

THESIS

Presented to the Graduate Council of  
Southwest Texas State University  
in Partial Fulfillment of  
the Requirements

For the Degree of  
MASTER OF PUBLIC ADMINISTRATION

by

Wilma Jean Cloyd  
(San Antonio, Texas)

San Marcos, Texas

May, 1980

TABLE OF CONTENTS

	PAGE
INTRODUCTION. . . . .	1
CHAPTER	
I. Evaluation Research for Social Programs . . . . .	2
II. Evaluation Research Components. . . . .	18
III. Case Studies in the Mental Health Field . . . . .	59
IV. Comparative Cases and Comments. . . . .	75
V. Evaluation Research and Additional Dimensions . . . . .	85
CONCLUSION. . . . .	88
REFERENCES CITED. . . . .	90
BIBLIOGRAPHY. . . . .	96

## INTRODUCTION

People decide to undertake social research for many different reasons, acknowledged and unacknowledged. A primary motivator for this rapidly growing activity is the rise in pressure from the tax payer. The demand is for accountability in the many social service programs funded through governmental entities at all levels. Legislators from the local to the federal level are being asked to demonstrate the success or failure of projects in terms of time, money and expertise expended.

In response to this demand, evaluation research in the field of the social sciences is being done in an attempt to provide answers. There are many and varied approaches or methodologies for accomplishing this goal, but the questions to be answered are basically the same. What is the ultimate goal sought by the public and/or the policy makers? What specific objectives are involved? Are efficient research strategies being utilized? Which social indicators or other measurement tools are most effective in acquiring appropriate data? Are research findings serving some purpose in the formulation of social policies and project development?

Ideally, social evaluation research will seek to judge the merits and/or failures of an activity, measured against the expenditures of scarce resources, for the purpose of providing a better, more rewarding life for the population involved in a program and for our society as a whole.

## CHAPTER I

### Evaluation Research for Social Programs

#### Background

The purpose of evaluation in social research is to measure the effects of a program against the goals it set out to accomplish using various types of social indicators (Weiss, 1972, p. 4). It is a means of contributing to subsequent decision making and improving future programming. The comparison of effects with goals stresses contribution to subsequent decision making, and the improvement of future programming denotes the social purpose of evaluation.

Research of any kind does not begin as a set of answers to specific human problems (Havelock, 1974, p. 42). It starts as a set of facts and theories about the nature of the universe, knowledge which can only be made useful to men through an extensive process of development. It has become so much an "article of faith" in the United States that basic science is useful to everyone that it is taught in our schools. It is assumed that our great advances have been due to scientific progress.

While most people have only a vague understanding of how new knowledge is transformed into something useful, there is a firm belief that scientific research can, and usually will, provide the answers to most of our societal problems. A visible indication of this assumption is the outpouring of funds expended for scientific research and related projects.

In 1976 the federal government invested more than \$1.8 million in knowledge production and utilization relating to the identification and solution of social problems (Lynn, 1977, p. 64). Included in this figure is research in the realm of statistics, evaluations, demonstrations and experiments. Though the need for large-scale support for social research and development is widely accepted, questions concerning its relevance to the making of social policy have become more insistent in recent years.

The beginning of systematic federal support for evaluation research can perhaps be traced to the creation of the Federal Bureau of Ethnology in 1881 (Lynn, 1977). During the following four decades, motivated by the progressive era's concern for social problems and the need for scientific advice generated by World War I, federal support for social research emerged in recognizable form. The depression era's social problems and World War II further stimulated federal spending for social research, which reached a level of \$53 million in 1937 and exceeded \$60 million by 1953. Growth was slow during the 1950's when our society concentrated on problems of national security and health.

In 1953 Festinger and Katz (p.v) published their theory that scientific methodology can be applied to human problems. This concept revolutionized psychology and seriously affected all branches of the social sciences. This stimulated another burst of governmental energy to solve social problems (Lynn, 1977). It accelerated sharply during the 1960's, with the present levels of spending being reached in the last few years.

The widespread unrest and dissatisfaction of the 1960's (Johnson, 1970, p. 2), generated massive and diffuse efforts to alleviate the

problems of society. These were compounded by a wider range of alternatives than ever before available, provided by a burgeoning technology. Attention was increasingly focused on evaluation research as a tool for gathering the kinds of information needed to determine whether any progress was being made.

Since the mid-60's (Bell, Nguyen, Warheit, & Buhl, 1978, p. 253), observers of health and human service systems have urged that public programs within a given geographic or social area be more directly guided by information about that area's needs. These observers include medical and social scientists, public officials, and other public program administrators who are engaged in policy formulation, planning, and program evaluation at all levels.

In the past decade (Tripodi, Fellin, & Epstein, 1978, pp. 1-2), administrators and planners have given considerable attention to the modification of existing public programs and to the future development of new programs to meet the needs of selected segments of the population. As more attention has been focused on public responsibility for social programs, demands for evaluation have proliferated. Program directors are now being asked by a more sophisticated general public to demonstrate not only the needs to which their programs are addressed, but also the contributions they make in solving or alleviating social programs.

With this increased emphasis on the need for program evaluation and the difficulties in implementing evaluations of social programs, numerous solutions have been offered. In an article succinctly entitled, "Let's Quit Stalling and Do Program Evaluation," Edwards and Yarvis (1977, p. 205) suggest workable solutions to these problems based on their own projects. Their purpose is to demonstrate concretely that program

evaluation is simple, feasible, useful, and practical. In their view, it is also relatively inexpensive when properly conceived and executed.

Program evaluation has become a very popular activity for social scientists. This is due in part to much recent social legislation which includes a requirement for evaluation of the legislative program (Riecken, 1976, p. 37). The wave of domestic social reforms in the 1960's that led to compensatory education, community action programs, manpower training, and measures for diminishing racial segregation and sexual discrimination has been responsible for the creation of a mini-industry of evaluation.

It has been said (Hoshino, 1978, p. 302) that the 1960's were a decade of analysis of social welfare programs, and that the 1970's will be a decade of evaluation. That is, there will be a shift from asking, "What happened?" to asking, "How good (or effective) is what happened?"

In addition (Tripodi et al, 1978), questions about the management of funds and the relative costs and efficiency of alternative programs are constantly being raised. For the administrator, the availability, appropriateness, and adequacy of program evaluation can determine the success and/or survival of an on-going program. For those applying for federal or state grants, the inclusion of a valid evaluation element may influence its appearance of feasibility.

Evaluation research is quickly achieving greater maturity as a discipline (Guttentage & Saar, Vol. 2, 1978, p. 11). Although it is still the newest social science field, developments within it have been quite rapid. Considerable consensual validation now exists about the domain of evaluation research and its distinctive features, but there is still great variety as to an exact definition. This will be addressed later in

this paper. The important factor is that the discipline, with various formats, is being used in every sector of society. The general thrust of these efforts is to assess the impact of public programs on many of the social, economic, psychological and political aspects of our lives.

Based primarily in the United States (Hoole, 1978, p. 16), evaluation research is being done by political scientists, sociologists, psychologists, educators, medical doctors, agricultural specialists, economists, anthropologists, lawyers, statisticians, and operations researchers, as well as legislators and bureaucrats.

The results of evaluation research studies are being published in numerous professional periodicals, journals, and books. There is also a new yearbook, Evaluation Studies Review Annual (Glass, 1976, Vol. 1; Guttentag & Saar, Vol. 2, 1977). Publication of the massive Handbook of Evaluation Research (Struening & Guttentag, 1975) clearly indicates that the field has come of age; the accumulated knowledge requires two volumes, 38 chapters, and 1,432 pages.

There is a recent trend toward international utilization of evaluation research. Such agencies as the World Bank, the International Labor Organization, The World Health Organization, the United States Agency for International Development, and the United Nations Educational, Scientific and Cultural Organization have undertaken special studies on evaluation research in the development field (Hoole, 1978).

#### Definitions of Evaluation Research

As previously mentioned, there is no commonly or uniformly accepted definition of the concept of evaluation research. Attkisson and Broskowski (Attkisson, Hargreaves, & Horowitz, 1978, p. 24) propose the

following working definition for program evaluation:

1. A process of making reasonable judgements about program effort, effectiveness, efficiency and adequacy,
2. based on systematic data collection and analysis,
3. designed for use in program management, external accountability, and future planning,
4. focuses especially on accessibility, acceptability, awareness, availability, comprehensiveness, continuity, integration, and cost of services.

This definition emphasizes the systemic role of evaluation in that it focuses on the utility and efficiency of the human service network, both in terms of specific client need(s) and the range of human service needs within communities (Attkisson et al, 1978).

Another good "working definition" of evaluation from Edwards and Yarvis (1977, pp. 205-206) is: "Evaluations are efforts conducted to determine what has happened." As evaluations become more sophisticated, they also provide answers to the question, "How much happened?" and then to the question, "How much happened compared to something else?" These questions are almost an echo to those asked by Hoshino (1978).

The problems of evaluation efforts are compounded by failure to define terms operationally, according to Deming (1975, p. 55), as well as by failure to lay down criteria by which to weigh gains and advantages against losses and disadvantages. Deming believes that any adjective which is to be used in evaluation requires an operational definition which can be stated only in statistical terms.

Rosener (1978, p. 457) has a less restrictive definition. She sees evaluation research as nothing more than the application of certain

kinds of research methods to the evaluation of social programs. She describes its purpose in terms almost identical to those of Weiss, i.e. "Its purpose is to measure the effects of a program against the goals it sets out to accomplish as a means of contributing to subsequent decision making about the program." Rosener also notes that evaluation research is a "Scientific" process which attempts to control as much as possible for the intrinsic subjectivity of the evaluative process. It does not purport to eliminate subjectivity, but rather to acknowledge it, and correct for it as much as possible.

Wholey (1976, p. 680) of the Urban Institute defines evaluation as "systematic measures and comparisons to provide specific information on program results for use in policy or management decisions." He observes that there are some evaluation processes which are relatively simple, inexpensive, and sufficiently timely to serve this purpose. He considers these to be the exceptions, however, rather than the usual situation.

Freeman (1974, pp. 15-16) maintains that evaluation research means more than studies of the utilization and impact of programs in the human resources field. He sees it as an emerging growth industry in which universities, nonprofit corporations and the profit-making sector are competing for grants and contracts. Millions of dollars are involved, but the stakes for the population to be served are high in another way. The past three decades have seen many social movements develop and then decline with varying degrees of success or failure.

The important point at this time is that interest and support for this expanding field of the social sciences has not diminished because of differences in scope or emphasis by individuals, groups or disciplines. Freeman (1974) believes that this is not so much a matter of who is doing what, but because of a fundamental belief that the utilization

of social science and social research techniques are the underpinnings of program development and resource allocation in the broad human resources field.

A few authors perceive a somewhat different purpose for social research. They focus on the application of systematic and appropriate knowledge in human affairs for the purpose of creating intelligent action and change (Bennis, Benne, & Chin, 1969, p. 4). The goal is planned change; a conscious, deliberate, and collaborative effort to improve the operations of a human system, whether it be self-system, social system, or cultural system, through the utilization of scientific knowledge.

The listing above of the several definitions of evaluation research in the social sciences is only one example of the wide variance, as well as the compatibility, among policy makers, social scientists and social practitioners as to the proper interpretation, direction and management of social change. Many people have come to believe that they have no actual choice as to whether others will be seeking to change the patterns of their lives. The only viable option seems to be to try to plan our changing futures to mesh as comfortably as possible with constantly changing technical and cultural conditions. The insights and services of behavioral scientists are being sought in some quarters, albeit belatedly, to identify social indicators more accurately, and to suggest procedures for anticipating and coping with prospective social changes.

As acceleration of the rate of technological and social change grows more and more rapidly (Bunker, 1978, p. 223), the complexity and interdependence of the socio-political and techno-physical environments also become greater and greater. Many factors must be assessed and choices

made. With rising demands for both equity and technical effectiveness in public programs, pressure is further intensified to utilize the contributions of science more systematically and effectively.

In most fields in which government applies scientific technology as an action tool, the use of scientific knowledge has been expanding in a natural, non-controversial, and largely successful way for some time. In contrast (Bunker, 1978), when there is input from the field of social science, a variety of problems reduce its influence and potential value. Factors such as the perceptions of validity of findings, theoretical propositions and the soundness of a proposal are frequently suspect. Whether social science knowledge will be used effectively or not depends upon the mind-set of decision makers about social science in relation to social action.

Bunker acknowledges that erroneous conceptions of social science among policy makers can distort or disrupt the utilization process, but admits that their remediation alone will not assure the linkage of policy content with social science knowledge. Changes will be needed, he feels, in both organizational structures and procedures.

Bunker points out the problem:

The government is prevented from learning from experience because it has no adequate and organized way of recalling what it has tried and what were the results. Similarly it has no systematic and functional way of finding out what questions it has asked and what the answers were (p. 224).

#### Social Indicators as a Measurement Tool

For over a decade, considerable research effort has been devoted to

the development of "social indicators." Denoting indexes of social life, this term was first introduced into the literature by Bauer in 1966 (Land & Felson, 1977, p. 328). Subsequent interest in social indicators has been so strong that it has been called a "social movement." Regardless of the perspective used to identify social indicators, the development of models for determining the interrelationships of changes in social indicators, or social indicator models, is necessary both for the analysis of social changes and for the establishment of sound social policies.

In recent years social indicators have been proposed as a component of statistical measures of societal conditions (Branch, 1978, p. 49). However, their expression in numbers does not make them intrinsically more accurate or meaningful if their derivation is imprecise, if their significance is limited because they represent only one aspect of the condition, or because their usefulness depends on interrelationships with other elements which cannot be determined reliably.

There is considerable interest in social indicators for cities, metropolitan areas, states, and regions (Fox, 1974, p. 130). Although computational problems in handling social system models of cities and regions are far from trivial, the major limiting factors at present are deciding what variables to measure, how to measure them, and how to estimate causal or structural relationships among them.

In answer to these questions, Branch (1978) suggests several qualifying approaches for identifying social indicators or measurement standards. He writes:

What is needed is a quantitative or other specific statement of the components of evaluation research for social programs.

It should be demonstrably reliable, sufficiently precise to permit comparison over time for a single city and also among different jurisdictions, and acceptable as a basis for executive, legislative, and judicial decisions.

The Social Science Research Council issued its first Social Indicators Newsletter in March, 1973 (Fox, 1974, p. 4). In it is perhaps the most comprehensive definition of social indicators and their role in evaluation research for social programs:

What are social indicators? We take them to be statistical time series that measure changes in significant aspects of a society. This is a minimal definition but for present purposes we think it is a realistic one, in light of the variety of meanings currently imputed to the term. . .The social indicator expresses something about the composition, structure, or functioning of that society, and expresses it in quantitative terms that can be compared with similar measures in the past or future.

The social indicators approach to need assessment is based primarily on inferences of need generally drawn from descriptive social and demographic statistics found in public records and reports (Bell et al, 1978, p. 266). The underlying assumption of the approach is that estimates of the needs and social well-being of those in a community can be made by examining selected social and demographic descriptors that have been found to correlate highly with service utilization. In other words, some social indicators are accepted by Bell and his associates as "empirical predictors of need."

Basic to this social area analysis orientation is the notion of a community (e.g., a city) as a constellation of "natural areas" (Bell et al, 1978, p. 267). A "natural area" can be defined most simply as a unit within a community, identifiable on the basis of characteristics that set it apart from other units. Some of the variables most commonly used to identify natural areas are topographical features such as rivers, terrain, and land-use patterns; modal sociodemographic attributes of the residents including age, race, sex, ethnicity, income, education, occupation, and family patterns; population factors including distribution, density, mobility, and migration; the spatial arrangements and distribution of institutions; and indicators of health and social well-being such as infant mortality, other age-specific mortality rates, crime and arrest records, suicides, and the prevalence of alcohol and drug abuse.

#### Principles and Problems of Evaluation Research

Ideally, programs are always conceptualized with clearly defined goals and objectives based upon some of the needs indicated by problems or factors like those listed above. With these goals spelled out well in advance of implementation, the social indicators carefully chosen, and the evaluation strategy identified with its components distinctly defined, a program is ready to proceed. The collected input data is constantly monitored to allow adjustments if indicated; concurrently the data is carefully recorded for evaluation at the conclusion of the program. The perfect "happy ending" includes accurate reports of the collected data, possible implications, questions raised, needs observed, and other conclusions written in a concise, interesting manner. Finally,

every attempt is made to place the evaluation research in the proper hands for appropriate utilization. This portrays the ideal. In the world of reality, the ideal rarely happens.

In Practical Program Evaluation for State and Local Government Officials (Hatry et al, 1973, pp. 20-22), the authors ask the question, "Why the current lack of comprehensive evaluations?" They list several understandable reasons, e.g. lack of funds and adequately skilled staff, controversial programs, and poor track records of quality and timeliness of evaluations. Hatry and his associates admit that their report does not examine "in depth" techniques, but focuses on approaches which are relatively inexpensive but may provide meaningful information adequate for most state or local government purposes. They write:

It is probably not necessary that evaluations conducted for local or even state governments conform to the stringent standards of professional social experimenters and evaluators. . .the information received from sound, though abbreviated, program evaluations can provide officials with valuable help for decision making (Hatry et al, 1973).

It is my impression that almost all social science writers and practitioners recognize the importance of every component of evaluation research, and that any missing part makes the evaluation inadmissible as scientifically pure. There are instances, it appears, in which half of a loaf is better than none.

According to Edwards and Yarvis (1977, p. 207), one should not assume that evaluations cannot occur because programs are started without clearly formulated goals and objectives. Usually, or in some cases,

if an evaluator waits for the provision of detailed goals and objectives from others, the evaluation never begins. A better alternative is for the evaluator to create measurable indicators which seem to be reasonable from the factors which are known. These may not be the most accurate or relevant, but they can start a dialogue between evaluators and administrators and allow initial data collection. As the evaluation evolves over time, the statements of goals and objectives will also evolve over time.

Edwards and Yarvis (1977, p. 210) use the term "endpoints" to mean proof positive of the impact of the intervention in a given situation. Once started, the evaluation process will evolve with unique relevance for a specific program. Success through demonstrating what has happened, and growing success through changing current practice, are possible only because evaluations are being conducted. They have found that many complex issues resolve themselves when evaluations actually are made within a program.

In addressing a group of architects, Raymond A. Bauer (1968, p. 237) indicated that it is necessary to understand our present state and its relationship to the past if an attempt is to be made to shape the future. This is the basis for continuing efforts toward understanding the trends now in effect; however, the data on many important aspects of our present society are missing, inadequate, or misleading. These flaws have long been recognized, but it now appears that greater efforts are being made to identify these problems.

At all levels of government, arguments are being made for improvements in the efficiency and effectiveness of service delivery (Scioli, 1979, p. 41). These arguments invariably translate into a plea for the

expansion of attention to policy evaluation. Ironically, the very condition demanding program evaluation is the same one negating it--scarce resources. Scarce resources demand that critical decisions be made regarding service delivery and cost effectiveness; yet the very data gathering, analysis activities and trained personnel needed are considered too costly to be regular program features. Thus the most general and pervasive problem surrounding program evaluation is this scarcity of resources. The major scarce resource in most cases is personnel who are qualified and experienced in the administration of evaluation research.

Social scientists with an orientation toward applied research have long been interested in evaluation (Brownstein, 1976, p. 1). A traditional topic was in the delivery of education and social services. Recently there has been an expansion of concern about evaluative research in "social action" programs in general and ameliorative programs in particular. Of particular interest is the fact that this research has taken on a generally interdisciplinary patina. As teams of individuals, with both strong personal perspectives and an appreciation of the need for the contribution of other areas of expertise, they have joined in searching for policy solutions to social problems.

Looking to the future, Prather and Gibson (1977, p. 563) point out a quote from Martinson (1975, p. 180) which states:

Future research must combine the analytical skills of the economist, the jurisprudence of legal advocacy, the sociology of the life span, and the analysis of systems. "Traditional evaluation" will play a modest but declining role.

Using research which is more systematic, together with increasing agreement on terminologies and their definitions, will hopefully bring about greater understanding of our basic social problems. This step forward should generate more accurate knowledge from which more efficient programs can be formulated.

Social indicators, as an adjunct to the many evaluative strategies now developed or being refined, can utilize electronic data processing to provide great quantities of information. With this huge mass of information immediately available, greater concentration can shift, in many cases, from data collection to the refinement and implementation of social programs to meet the needs of a given population or an entire nation.

Bauer (1968) states that "what is needed is a little sociology, a little statistics, and a lot of brains."

## CHAPTER II

### Evaluation Research Components

#### The Goals of Evaluation

One of the reasons for the failure of evaluation to influence decision-making is that the goals of evaluation have often not been clear. An evaluator may be called in to study the effects of a program and not be told its purpose (Weiss, 1972, p. 34). If he presses for a statement of goals, program administrators may answer in terms of the number of people they intend to serve, the kinds of service they will offer, the types of staff they will have, and similar information. For program implementors, these are "program goals" in a real and valid sense, but they are not the primary currency in which the evaluator deals. He is interested in the intended consequences of the program in clear, specific, and measurable terms, rather than the program peoples' fuzzy replies of trying to "improve education" or "enhance the quality of life." Evaluation can have many different goals. For example:

1. Feedback for program personnel.
2. Evaluation of the performance of individual personnel.
3. Evaluation of the quality of programs.
4. Providing accountability to the community or clientele.
5. Increasing the efficiency of the organization.
6. Providing justification for the funding of the organization.

There are many types of purposes for which evaluation research can be used. For evaluation purposes, broad or generalized ultimate goals of

a program must be refined into a list of specific objectives. There objectives must then be translated into operational terms, make visible, and measurable. This is not as restrictive as it may appear.

Epstein and his associates (1977) suggest a simple devise for translating abstract goals into specific objectives. Identify who in the program (which agency personnel) are expected to do what (using what procedures, strategies, therapeutic approaches, provision of services, etc.) to whom (the designated target population), where (at what location), when (with what frequency), and why (to accomplish what ends or objectives)? Once these questions are answered, criteria can be developed for these specific objectives.

#### Criteria for Evaluation Objectives

The basic question (Bauer, 1966, p. 39) is, "What is to be measured?" This question can be answered only if a more fundamental question is asked and answered, "For what purpose?" Obviously, the criteria for evaluating a program of educational reform or a program for stimulating productivity in industry will be different.

Program objectives are the ends to what program efforts are directed (Epstein & Tripodi, 1977, p. 119). The term "program" refers to an activity or group of activities undertaken by a government to provide a service to the public. It may be contained within a single agency, but more typically it involves the meshing of efforts by various parts of government and private agencies.

According to Hatry, Winnie, and Fisk (1973, p. 8) program evaluation research concentrates on identifying how the condition of citizens and the community have changed as a result of a specific program or set of activities. The major criterion is the attempt to determine whether a

program is achieving government objectives, and then to consider both its positive and negative impacts. Consequently, the evaluation helps policy officials determine whether a program should be continued as is, expanded, modified, reduced, or eliminated. If a program is not performing as expected, the evaluation may help indicate reasons for ineffectiveness and action which may be taken to remedy the situation.

More than one objective and evaluation criterion needs to be considered (Hatry et al, 1973). Inevitably a program involves numerous objectives, and numerous evaluation criteria will be needed to measure their effects. Objectives and evaluation criteria should cover each negative effect.

Evaluation criteria should not be rejected because of apparent difficulties in measuring them (Hatry et al, 1973). There are often ways to at least partially measure the more qualitative, subjective types of evaluation criteria such as use of ratings, rankings, and other procedures which can be quite useful and appropriate if undertaken in a systematic way.

Hatry and his associates believe that it is probably better to err on the side of too many objectives, evaluation criteria, or clientele groups than to eliminate some that might be important when examined more closely. Neither public officials nor program evaluators should eliminate a potential evaluation criterion on the basis of their own personal observation. Program effects on individual population groups should be distinguished since various groups may be affected by a program in different degrees.

Dollar cost should always be included as a program criterion because governmental units will always want to know what an on-going

program has cost as it considers future projects. Past costs can be compared to effects and these figures may be used to derive estimates of likely future costs if the program is continued. This is a very complex procedure, but it is one which cannot be ignored (Hatry et al, 1973).

Goldenberg (1979, p. 94) critiques four books whose themes are evaluating municipal services. He notes that the authors of each of these books apply certain criteria to judge government performance. These criteria are more or less explicitly stated, but they vary tremendously across the studies. The lack of agreement on the proper criteria for evaluating governmental services is characteristic of the field as a whole. It is also typical to gear an entire analysis to only one criterion and ignore the others.

#### Evaluation Research Design

Wrightsmann (1977, pp. 36-39) describes the scientific research method as a cyclical process of establishing facts. Its main steps are: (a) induction of one or more theoretical propositions from observed facts, (b) deduction of some logical consequences of these theoretical propositions, and (c) verification of the predicted consequences by collecting new observations. He suggests four basic ways of studying human behavior:

1. Collecting and analyzing existing products.
2. Asking questions.
3. Watching people.
4. Manipulating conditions experimentally.

Within these basic approaches, he lists many of the major research methods or designs (Wrightsmann, 1977, p. 45). They are: (a) archival

research, (b) surveys, (c) field studies, (d) natural experiments, (e) quasi-experimental research, (f) field experiments, (g) simulations, and (h) laboratory experiments. He concedes that there are problems inherent in each of these basic research approaches.

The entire research process can be thought of as a continuous development (Wechsler et al, 1976), in which each decision influences the next. Thus, the manner in which a problem is formulated in turn sets the stage for the next step, the choice of a design. The research design is basically the outline or blueprint of the study. It should follow logically from the particular manner in which the problem is framed. The most important factors in the ultimate decision on what type of design should be used are the present level of knowledge about the program, group, or phenomena to be studied, and concomitantly, the degree to which one wishes to be able to generalize the findings of a particular study to encompass a broader group of people, agencies, or programs.

The very purpose of research (Wechsler et al, 1976), to contribute to knowledge, to answer questions, almost implies that original data, data collected anew to help to answer the research questions, is needed. It seems evident that the data requirements of the current emphasis on matters such as accountability, program evaluation, and hypothesis testing research cannot be met through established data sources. In many research efforts numerous sources are used for the gathering of original data.

Exploratory studies focus heavily on ways of collecting new data relevant to the subject of the research. Descriptive studies frequently use concepts drawn from the social and behavioral sciences. Examples are social class, family structure, social functioning, and self-concept.

In order to classify or categorize phenomena on the basis of these formulations, original data must usually be collected on the group under study. The central role of original data in experimental design is self-evident (Wechsler et al, 1976).

In practice, the nature of the research problem, the level of the study, and the design to be used tend to indicate which sources of data are relevant. The choice of procedures will determine the cost and manpower requirements of the measurement process as well as the data's accuracy, and this affects whether the information obtained is worth its cost. To lower costs and increase credibility, it is often wise to make maximum use of data already collected. However, in many cases new data is vital.

Discussions of evaluation research designs are prolific, but, with a few additions and/or refinements, most evaluations fit into five basic designs. Hatry, Winnie, and Fisk (1973, p. 39) elect to call this major step in program evaluation design "Comparison--The name of the Game."

They write:

Ideally, we would like to compare what "actually happened" to what "would have happened if the world had been exactly the same as it was except that the program had not been implemented." Since it is impossible to determine exactly what "would have happened if . . .," the problem is to use procedures that approximate this.

Their approach for identifying and quantifying program effects due to the program are:

1. Before vs. after program comparison--compares program results from the same jurisdiction measured at two points in time: immediately

before the program was implemented and at some appropriate time after implementation. They consider this to be the "bargain basement evaluation."

2. Time trend projection of pre-program data vs. actual post-program data--compares actual post-program data to estimate data projected from a number of time periods prior to the program.

3. Comparisons with jurisdictions or population segments not served by the program--compares data from the jurisdiction where the program is operating with data from other jurisdiction where the program is not operating.

4. Controlled experimentation--compares pre-selected, similar groups, some of whom are served and some of whom are not (or are served in different ways). The critical aspect is that the comparison groups are pre-assigned before program implementation so that the groups are as similar as possible except for the program treatment. This is considered to be "the cadillac of program evaluations."

5. Comparisons of planned vs. actual performance--compares actual post-program data to targets set in prior years--either before program implementation or at any period since implementation (Hatry et al, 1973).

Selection of an evaluation design depends upon the timing of the evaluation (if the evaluation is decided upon after the program has already been implemented, design four will not be possible), the dollars available, and the accuracy desired.

The first four designs discussed above are progressively more expensive (Hatry et al, 1973), with the fourth usually considerably more expensive than the others. The first three and the fifth can often be accomplished with but a very few man-months of analytical effort--the

amount depending heavily on the amount of special data collection required. The fourth design will likely take many calendar months and possibly years. The designs are also progressively more effective in providing substantial evaluation information; the fourth providing, by far, the most reliable results. The designs are not either/or choices. Some or all of the first three are often used together.

#### Social Measurements: Techniques and Indicators

Once goal statements are clear and unambiguous, skilled researchers can measure all manner of things (Weiss, 1972). They can use the whole arsenal of research techniques--observation, content analysis of documents, testing, search of existing records, interviews, questionnaires, sociometric choices, laboratory experiments, game playing, physical examinations, measurement of physical evidence, and so on. With attitude tests and opinion pools, they can measure even such relatively "soft" goals as improvements in self-esteem or self-reliance. The evaluator will also want to find and measure the behavioral consequences of changed attitudes--the things participants do because they feel different about themselves, other people, or the situation.

The significant factor for the researcher is the development of indicators to measure the extent to which the goals are achieved. These indicators of program outcomes are the dependent variables of the study. Also important is the description and measurement of other relevant factors such as the inputs, which are the independent variables of the study, and possibly intervening variables, factors that mediate between inputs and outcomes.

The development of measures is a demanding phase of evaluation. If the evaluator is in luck, existing measures which are suitable for the

subject of his concern may be located (Weiss, 1972). It is worthwhile to search for these measures that have already proved workable rather than to create new ones. It may eliminate much of the trial and error work. Also, repeated use of common measures helps to build up a body of knowledge, and it becomes possible to begin to make comparisons about the relative factors under investigation.

Measures that are useful for assessing outcomes in evaluation research (Weiss, 1972) depend on the research objectives or program intent. They deal with attitudes, values, knowledge, behavior, budgetary allocations, agency service patterns, productivity, and many other items relevant to the problems being studied. A knowledgeable evaluator is invaluable at this point. He is aware of past research programs, the types of indicators and measures used, the most suitable research design for the type of program, as well as the most advantageous usage of the resources of time, personnel, and money available. It is unfortunate that such trained personnel are among our most scarce resources.

Etzioni and Lehman (1969, p. 62) make a fairly complete summary of the traps which any social quantification may conceal when it is carried out with insufficient strictness. Errors may arise from three sources: (a) the fractional character of the measurement, (b) its indirect character, and (c) the special difficulties of measuring collective attributes.

There has been a movement to develop a national system of social indicators. This would institute common categories, terminologies, definitions, and scope of data so that these items can be more accurately utilized for comparisons, especially in time-series studies. This will not solve all of the evaluator's problems, in any sense, but the very

word "indicator" suggests its limitations. It will allow better social accounting, and it is an accepted factor of any social research that there is no way of separating out the effects of a program from all of the other factors operating on the indicators.

Etzioni and Lehman (1969) point out that since all social measurements face the problem of internal validity, persons who use these measurements, either for research or for policy making, ought to be alert to dysfunctions which may emerge. Increased investment, intellectual as well as financial, no doubt can go a long way to increase the efficacy of social measurements and to reduce much of the likelihood of dysfunctions. But, in the final analysis, these problems can never be eliminated entirely. Hence, the client of systematic measurement and social accounting should be alerted to the limitations of social indicators, both to make his use of them more sophisticated and to prevent him from ultimately rejecting the idea of social accounting when he encounters its limitations.

Weiss (1972) regards the choice of indicators for evaluation as critical. For massive programs, such as public housing or Medicare, social indicators can provide time-series data on the distribution of resources and outcomes. They can have the advantage for federal decision purposes of using common criteria and collecting comparable data across projects and across time, and if astutely constructed, dealing in issues of relevance to policy makers. She also notes that social indicators cannot overcome such inherent limitations as the failure to account for external influences or the absence of information on causes and dynamics of change. But if supplemented by, and related to, specifically evaluative studies on critical issues, their information on nation-wide conditions can be supportive and important.

Other social scientists (Bell et al, 1978) see some major advantages in using the social indicators approach. First, social indicators can be developed from vast data pools already existing in the public domain, e.g., census reports, governmental agency statistics, and social research organizations. Second, most social indicators can be secured at relatively low cost by persons with a limited amount of research training or technical expertise. Third, a social indicators analysis can either include only a selected amount of information about a local community or include comparable data about other communities and the country at large. It is also possible to integrate data to sociodemographic characteristics, social behavior and well-being, and community conditions.

Bell and his associates (1978) take the position that multiple sources of data permit agencies with differing levels of resources to select the data analysis procedures that are suited to their objectives and commensurate with their resources. Fourth, the results of a well-executed social indicators analysis can serve as a foundation on which to validate or update other need assessment approaches with a minimum of effort and cost. And fifth, because social indicator data are collected uniformly across the United States, it becomes possible to use these data to compare communities or to control for demographic differences when comparing health and human service agencies. Altogether, social indicators approaches have a great deal to commend them.

Most of the disadvantages in using social indicators to assess needs (Bell et al, 1978) are theoretical in nature. The first and most serious drawback arises from the fact that many of the social indicators are only indirect measures of the needs they are supposed to represent. Consequently, there is reason to question the validity of some social

indicators. For example, divorce, separation and illegitimacy rates are often used as indices of family stability and security. However, these factors do not represent family instability for all groups in our society. Second, users of the social indicators approach have been criticized for their tendency toward ecological determinism. Although the spatial characteristics of an area obviously influence and limit the social conditions and relationships in the area, these characteristics do not totally determine or, in an absolute sense, cause the social conditions.

Third, there is the danger of overgeneralization of the findings of a social indicators analysis. Taken to extreme, this practice of drawing inferences at the individual level from data and relationships observed at the social area level is equivalent to stereotyping. It can be misleading, counterproductive, and potentially damaging (Bell et al, 1978).

However, these theoretical issues and drawbacks should not dissuade the use of social indicators to assess health and human service needs, according to Bell. The goal of social indicators studies is to analyze the relationship of spacial, sociodemographic, and social conditions of an area with the social behavior and well-being of the area's residents.

Weiss (1972, p. 59) suggests that perhaps the gravest impediment to the use of social indicators for evaluation is that it expects so much. A little bit of change is not enough; programs have to move from having people "unemployed" to "employed," and situations have to change from "massive air pollution" to "approved air standards."

Even if change does come (Weiss, 1972), it often appears slowly. Indicators are sluggish. They are derived from periodic soundings, usually annual, so that there is a considerable time lapse before trends become apparent. By the time changes appear in the figures, numbers of

other influences have been operating on conditions, and one is back to the problem of authenticating the program as the source of the effects.

Regardless of the specific measurement process, it is concerned with finding an expression for the degree of difference in distinguishable qualities or characteristics (Lichfield, Kettle, & Whitebread, 1975, pp. 98-99). The "why" of measurement is that more information can be obtained than otherwise could be the case and this increase facilitates the decision making process.

Lichfield and his associates list three purposes for measures. They are: (a) to facilitate comparative descriptions through the indication of differences, (b) to assist in acquiring knowledge about certain phenomena, and (c) to formally test propositions and theories. The main justification for measurement in evaluation is that it reduces the extent of subject judgement, although, as most writers agree, this can never be eliminated because uncertainty inevitably surrounds the quantification of forward-looking estimates.

The choice of measurement procedures will determine both the financial and man-power requirements, and this affects the data's accuracy and whether the information obtained is worth its cost. To lower costs and increase credibility, it is often wise to make maximum use of data already collected; however, in many cases some current data is required.

It is very difficult to measure benefits in terms of general social well-being (Dye, 1972, p. 294). The movement to develop social indicators to measure the social well-being of American society is a step forward, but we are still a long way from assessing the impact of social programs. When some commonly accepted indicators with their definitions are more widely accepted, it should be possible to more rationally evaluate alternative public policy by weighing their costs against gains in social indicators.

### Data Collection

The assembling of information for research purposes may be distinguished from the more general collecting of impressions and observations by the nature of research itself (Wechsler, Reinherz, & Dobbin, 1976, pp. 139-142). Data gathering is a disciplined, systematic endeavor to examine specified phenomena in an objective, measurable manner. Embodied in the formulation of the problem and elaborated in the research design are the crucial elements or variables that point to the appropriate data fields for investigation, and specify which items of data are relevant to a given study. These items then become measurable indicators.

Two general sources of data may be readily identified (Wechsler et al, 1976). One is the data on hand, data in a variety of places, already collected and presumably available to the researcher. These existing collections of data are sometimes referred to as secondary data sources or available data sources. The second general source of data may be thought of as the actual phenomenon under study.

Extensive automated data banks in many fields of human service provide large amounts of valuable data economically. Perhaps the most widely used are the data files developed by the Bureau of the Census. A wide range of useful information is also collected and stored periodically by local, state, and federal government agencies (Wechsler et al, 1976).

In deciding on a research strategy for data collection (Wechsler et al, 1976), the researcher must first be aware of the existing knowledge base. In this sense, all research may be said to be based on former research. In formulating the research problem, the practitioner is responsible for using what is known, the available data sources, as a frame of reference. Thus we have the familiar review of the literature in

research, and the very useful state of the art reports produced in recent years.

Particular strategies for gathering research data include the abstraction of data from secondary sources, observation, and techniques such as interviews, questionnaires, and a variety of test and measurement instruments.

Tests and measurements, drawn from the field of psychology, provide another method and source for collecting useful data (Wechsler et al, 1976). The strategy is to collect data through indirect approaches such as projective techniques. The underlying assumption is that through well-developed techniques, the necessary data can be collected to define, describe, quantify, and qualify such illusive phenomena as feelings, beliefs, attitudes, personal orientations, and basic characteristics.

Familiar approaches used include the Rorschach, Thematic Apperception Test and Self-Concept scales. Indirect methods rely heavily on "proxy" variables. They draw data on responses to issues and situations that reflect or are assumed to stand for what the researcher is attempting to measure (Wechsler et al, 1976).

The physical components of a city can be counted or measured and the condition evaluated by visual examination, engineering, scientific analysis, or aerial remote sensing (Brnach, 1978, pp. 53-54). The data provided by such methods are accurate enough to support conclusions and actions relating to the quantity and condition of many physical elements and aspects of city planning.

The socioeconomic state of the city which generates its physical state and most directly represents the condition and contentment of its population is more difficult to determine. Economic status can be gauged

through income data, but these data do not take many factors into account such as debts and other obligations. Income data are statistically less reliable than generally presumed, particularly with respect to minority groups and the disadvantaged for whom such indicators are most needed.

Living conditions are not easily evaluated and personal contentment with living conditions and satisfying relationships with other people are probably the most important indicators of the socioeconomic situation of each person. These factors are the most difficult to measure individually and collectively, but such data would disclose increasing social tensions and possibly anticipate potential disturbances. Experiments need to be developed which will establish beyond a reasonable doubt the significance of the unconscious effects of the urban environment on its inhabitants, and measurements devised to indicate the severity of these effects (Branch, 1978).

Another important procedure in collecting data can be labelled the "trained observer" approach. Such procedures must be systematic enough for different observers to give approximately the same ratings when observing the same conditions. One such rating procedure includes photographs representing different levels such as cleanliness, numbers of users, or other conditions. Using them as a rating guide, different inspectors, at different times, and in different parts of the city, can give approximately the same ratings for similar conditions--even the same observer can use the photographs as reminders (Measuring the Effectiveness of Basic Municipal Services, 1974).

In selecting measures of effectiveness in data collection procedures, Measuring the Effectiveness of Basic Municipal Services: Initial Report (1974), recommends that the following criteria be considered:

1. Appropriateness and validity. Does the measure relate to the government objective for that service and does it really measure the degree to which a citizen need or desire is being met?
2. Uniqueness. Does it measure some effectiveness characteristic that no other measure encompasses?
3. Completeness. Does the list cover all or at least most objectives?
4. Comprehensibility. Is the measure understandable?
5. Controllability. Is the condition measured at least partially the government's responsibility? Does the government have some control over it?
6. Cost. Are cost and man-power requirements for data collection reasonable?
7. Accuracy and reliability. Can sufficiently accurate and reliable information be obtained?

Interviewing former clients, both graduates and drop-outs, of treatment programs with hoped-for long-term impact (such as drug abuse, man-power training) can provide vital information on the effectiveness of these programs. Lack of such follow-up is a major weakness in current evaluation efforts, and it will be discussed at more length later in this paper.

Information gained through measurement provides a basis for evaluation when it is placed in the hands of the appropriate persons or groups of persons, and it is supposed to influence decisions. Abert and Kamrass (1974, pp. 37-38) observe that it is premature to judge how influential such evaluations have been in reshaping social policy, but experience to

date suggests that there are certain difficulties associated with the usual and ordinary procedure of conducting post-hoc evaluations of national programs, i.e., with the common administrative practice of waiting until the program has been put into full operation before giving appreciable attention to its evaluation. (This was probably the understatement of the 1974 social science year.)

Social programs have proliferated with seeming abandon, but of late the efficacy of such programs has been widely debated. Prather & Gibson (1977) believe that many purported evaluations of government are neither reliable nor valid. Perhaps Rivlin (1971) was correct when she stated "Little is known about how to produce more effective health, education or other social services. . . We simply do not know the best or most effective ways to produce effective [social] services."

Possible sources for identifying relevant measurement data are rarely found neatly packaged and described. However, in Hatry, Winnie, and Fisk's Practical Program Evaluation for State and Local Government Officials (1973), there is a list of a variety of sources which may provide important clues to what these are. The list includes:

1. Legislative statements relating to objectives and evaluation criteria. These are usually on state and federally originated programs--at best, partial sources.
2. Expressions made by legislators or citizens at hearings before a local council, in the press, or by organized citizen groups.
3. Program personnel will often be aware of many unintended as well as intended consequences, both

beneficial and negative, as well as the various population segments that appear to have been affected.

4. Government officials themselves. This may prove to be difficult because of political expediency, but if there is too much "hidden agenda" the evaluation is likely to lack utility and should not be attempted in the first place (p. 36).

A few other sources which should be listed are: (a) analysis of citizens' complaints to local governments, (b) citizens' concerns as expressed in a number of small newspapers, (c) professional literature, and (d) suggestions by reviewers of project draft reports (Measuring Effectiveness, 1974, p. 5).

Most organizations fail to take advantage of one of the most obvious sources of evaluative data, their own records (Rocheleau, 1975, pp. 39-41). They are required by law and/or regulatory agencies to collect huge amounts of data which are rarely used by program personnel. There have been recent attempts to make records more useful with a problem-oriented approach. Used originally in the field of medical practice, it has been extended to other areas which have a clinical or case approach, e.g., mental health, juvenile delinquency, welfare and related fields.

Most records are not organized around any systematic principle. Rocheleau (1975) describes the problem-oriented record system as being organized around the problems, plans, and goals indicated for each client. The major components are:

1. Data base--includes history of client, present illness or problems, and results of any examinations

and tests that have been carried out on him.

2. Problem list--includes both the complaints of the client and the analysis of the therapist or caseworker, comments by "significant others," and problems indicated by any tests performed on the client.
3. Plans and goals--a specific plan and goal is associated with each problem in the problem list.
4. Follow-up--includes progress notes and further feed-back concerning old problems or new problems which have emerged.

These four states are carried out in sequence with the feedback from follow-up being used to start the whole sequence over again with new data, new problems, and so forth.

Rocheleau (1975, p. 42) admits that it is difficult to summarize the advantages and disadvantages of this system because it is implemented in such a large variety of ways. However, he contends that past experience with them indicates they can be of value when attention is given to the following points:

1. The record system includes a regular follow-up.
2. Some measures of comparison such as levels of functioning is used to facilitate assessment of treatment outcome and comparisons among different therapists, programs, etc.
3. The records are routinely monitored for quality, reliability and validity.

Fitzpatrick (1970, pp. 74-75) views the many measures and criteria which might be proposed for an evaluation and asks how one can choose among them. From one point of view the answer is easy: apply the criteria of relevance, comprehensiveness, reliability, and feasibility, along with the professional judgment of the evaluator, making whatever compromises and innovations are necessary to produce a balanced and cost effective set of measures. But of course, if one has the skill to do all that, he is in little need of advise.

Guttentag, and Saar (1977, p. 12) concur that uninformed decision making can be costly to society. Currently the major thrust of evaluation research is the conceptual and methodological attempt to integrate and aggregate data so that realistic and timely inferences can be made from evaluative information. The challenge and responsibility for evaluation research is how to provide policy makers and the general public with the most accurate and relevant information. The emphasis has been on data production. Now, both the large amount of information naturally available from many programs and concerns about costs of information generation require evaluators to exercise caution about the use of resources for massive data generation operations. It is increasingly important to concentrate on the best utilization of data by incorporating it into the most appropriate evaluation research design.

Once the data have been collected (Epstein & Tripodi, 1977, p. 156), it is possible to determine the relative effectiveness of the intervention strategies being compared. In after-only studies, percentage comparisons, of mean scores, and so on are made between the experimental groups. Chi-squares may be computed to determine whether differences in the effectiveness scores of these groups are statistically significant. The groups

with effectiveness scores that are significantly higher have received the more effective intervention.

In before-after studies, comparisons are made within experimental groups over time as well as between experimental groups. Before-after comparisons within a group indicate the impact of the program on that group itself. Comparisons of post-intervention scores between experimental groups indicate the degree of success achieved by each given strategy. Comparisons of the before-after changes between experimental groups yield the most precise indication of the relative effectiveness of the various programs (Epstein & Tripodi, 1977).

After the effectiveness data have been analyzed, questions of relative efficiency should be raised. The costs of implementing the various interventions should be determined, and cost/effectiveness ratios should be computed for each intervention. Decisions about future program strategies should be based on relative effectiveness, relative efficiency, and, if possible, information about any undesirable side-effects of implementing specific programs (Epstein & Tripodi, 1977).

Prather and Gibson (1977, p. 557) are persuaded that the most basic factor of a reliable evaluation is the research design itself. In their view, evaluation reports concentrate on two basic types of research designs: descriptive and experimental. Descriptive designs tend to inundate the decision maker with data, but seldom answer such key questions as the program's comparative effectiveness, opportunity costs, consequences, etc. In agreement with Hatry and his associates, their basic rationale of the experimental approach is comparison, whether between alternative treatments, or between the treatment and random noise effects.

Many of the research designs employed in evaluating programs of social action are, at best, halfway adequate (Nunnally & Wilson, 1975, p. 229). Social action is often concerned with changing psychological attributes that are very difficult to measure at the present time. For example, one can count the number of children taught to read, but such subtle changes as aspects of attitudes or desire for knowledge, while not hopeless, are very difficult, to say the least.

There are arguments both for and against increasing attempts to make experimental studies and other types of research on human behavior relevant to the problems of our society (Wrightsmann, 1977). Despite this, the trend toward greater social relevance of research seems unmistakable.

A basic problem with experiments is that their very nature, most are artificial situations that are not intended to duplicate any situations in the real world (Wrightsmann, 1977, pp. 72-73). Even less are they intended to solve major social problems. However, if experiments seem far removed from real life, it must be remembered that any small bit of information which results adds to our basic store of knowledge and may prove to be part of the solution to a problem. As Festinger (1953, pp. 169-170) stated, "Experiments in the laboratory must derive their direction from studies of real-life situations, and results must continually be checked by studies of real-life situations."

Experimentation thus is not in itself viewed as a source of ideas necessarily contradictory to traditional wisdom (Campbell & Stanley, 1963). It is rather a refining process superimposed upon the probably valuable

cumulations of wise practice. Advocacy of an experimental science does not imply adopting a position incompatible with traditional wisdom.

In their classic monograph "Experimental and Quasi-Experimental Designs for Research," Campbell and Stanley (1963) identify eight factors that can jeopardize the internal validity of inferences drawn from a study. Epstein and Tripodi (1977) define each of them as they apply to the area of program evaluation. They are:

1. Contemporary history--unanticipated events may occur while a program is under way that change the character of the intervention, the client's situation, or the client himself.
2. Maturation--during the course of program intervention clients may change simply as a function of time, developmental growth, fatigue, and so on.
3. Initial measurement effects--the process of measurement itself might affect client outcomes.
4. Instrumentation--unreliability over time due to lack of standardization of the measure.
5. Statistical regression--the tendency of research groups selected for intervention on the basis of extreme scores on some index of need or pathology to "naturally" regress to a more average score in subsequent testing regardless of the effects of program interventions.
6. Selection--differences between experimental and control groups, or among groups receiving different kinds of interventions, can yield misleading findings.
7. Subject mortality--certain types of subjects may drop

out of the program in disproportionate numbers, creating mis-leading findings.

8. Interaction effects--the combined effects of any and all of the above factors may be mistaken for the effects of program interventions.

When the objects of research require that people be studied in their normal surroundings (Warwick & Lininger, 1975, p. 9-11), the choice of strategies usually comes down to survey research, participant observation, or some combination of the two. The first makes use of structured questions to a carefully controlled sample, while the second elicits information through the investigator's intensive, but less structured, interactions with a group. This comparison is particularly helpful in showing the uses and limitations of the sample survey, for its strengths are the prime weaknesses of participant observation, and vice versa.

The sampel survey (Warwick & Liniger, 1975) is an appropriate and useful means of gathering information under three conditions: (a) when the goals of the research call for quantitative data, (b) when the information sought is reasonably specific and familiar to the respondents, and (c) when the researcher himself has considerable prior knowledge of particular problems and the range of responses likely to emerge. All of these conditions are met in the areas of research that have been the traditional strongholds of the survey--public opinion, voting, attitudes and beliefs, and economic behavior.

Participant observation is usually more appropriate when the study requires an examination of complex social relationships or intricate patterns of interaction (Warwick & Lininger, 1975), such as: (a) kinship obligations or gift exchange in tribal villages, (b) when the investigator

desires first-hand behavioral information on certain social processes, such as leadership and influence in a small group, (c) when a major goal of the study is to construct a qualitative contextual picture of a certain situation or flow of events, and (d) when it is necessary to infer latent value patterns or belief systems from such behavior as ceremonial postures, gestures, dances, facial expressions, or subtle inflections of the voice.

Observation is widely used and is frequently the central data gathering strategy, especially in exploratory studies (Wechsler, 1976). Some studies concentrate on the direct observation with the use of a camera to capture data. The development of video technology and techniques in recent years has added new dimensions to this aspect of data gathering.

The number of observers (Wechsler, 1976), the roles they are to play, the focus of their concern, and the means they use to record observations, vary with the nature of the research task. However, much of the data essential to social research cannot be collected through observation, and for these situations other strategies have been developed for the gathering of data. Through the use of interviews and questionnaires, information is collected on phenomena such as peoples' attitudes, feelings, and past experiences. This approach is based on the person's own statements or written responses to inquiries. Such self-reports are widely used in social research. The possible pitfalls involved in these strategies center on the validity and reliability of the reports.

One difficulty is that many, if not most, national programs are not designed so as to permit comparison with alternative treatments (Abert & Kamrass, 1974). After-the-fact evaluation of a program's effect is intrinsically handicapped because the program was not designed in the first place to make available the systematic information needed to assess its

effectiveness. Such criteria can be designed into the administration of the program at its outset to assure explicit recognition of the need for deliberate and designed variation of the treatment for program assessment purposes.

Virtually every student of policy evaluation has acknowledged that the real impact of governmental programs may not appear until after a considerable period of time. However, very few research studies have included a time dimension as a factor to be measured (Salamon, 1976, p. 258). In his pioneering evaluation of the Volunteer Work Camps, Riecken (1952, p. 22) noted that "all experience with action programs indicates that their real effects cannot be gauged without considering the long-run forces that may support, negate, or even reverse the immediate effects." Despite the importance of the time dimension in evaluative research, however, the subject has so far received little more than passing mention. Thus, while there is general agreement that program impacts can decrease, increase, or remain the same with the passage of time, evaluation literature provides little theoretical guidance.

There are no suggested guidelines about which of the several results the evaluator may expect under what circumstances, or about what time perspective is appropriate for what type of programs (Salamon, 1976, pp. 265-278). To illustrate this fact, Salamon wrote of a detailed evaluation performed in 1973-1974 of a rural resettlement program for both black and white sharecroppers. The initial program began in 1934 as part of a New Deal program and it extended until 1943. The data which emerged showed that the land reform experiments had a strikingly significant sleeper effect, especially for the blacks. It created an important, black,

landed middle class, independent and confident enough to shoulder the burden of challenging the two-caste system once conditions became ripe.

Salamon concluded that:

Social change is a complex process, frequently moving in strange ways, and by fits and starts, to work its effects.

Evaluations of social action programs designed to promote such change must be equally resourceful. . . (1976, p. 281).

The search for more unobtrusive measures to replace or supplement our reliance on interviews and questionnaires is a welcome trend in evaluation research (Wrightsmann, 1977, p. 74). Also there is more research in natural settings and greater integration of field studies with laboratory studies, a systems approach to understanding human behavior, and a growing concern for protecting the rights of human subjects and ensuring careful adherence to ethical principles in research. The latter concern may have the unfortunate side effect of stifling much valuable research through restrictive legislation.

It should be acknowledged, however, that unrealistic expectations have often been raised about the results that can be achieved by evaluation research. Many programs are operated without a clear and agreed-upon statement of the objectives of the program, expressed in measurable terms; many programs operate on a very limited scale, providing resources to meet only a small fraction of the need; often evaluation studies seem to be inconclusive regarding the impact of the program; and, not infrequently, research studies are filed, unread, only to fade into oblivion. Evaluation research in and of itself does not perform feats of magic.

Our society has been using a system of economic indicators for some time (Bauer, 1968, p. 250). Now, the need for social indicators with

commonly accepted meanings is clearly apparent. In 1969 (Wrightsmen, 1977, p. 73) a presidential task force was appointed in the United States to develop social indicators to measure the state of citizen's morale, health, and well-being in the same way that the gross national product and the stock market have traditionally measured our economic condition.

Technical advances in electronic data processing have made the gathering, storing, and retrieval of information comparatively easy. The primary obstacles now are the determination of types, amounts, and ethical methods of using the information to better our society as a whole while protecting the privacy of the individual.

On the other hand, perhaps the greatest single obstacle to greater utilization of our advanced technology is that many people are inclined to be comfortable with the familiar and will not accept better answers. As Janis and Mann (1977, pp. 182-193) state it, "Psychological resistance to realistic information at any stage in the decision-making process results in errors in the decisional balance sheet."

#### Utilization of Findings

With all the possible uses for evaluation to serve, the evaluator has to make choices. The all-purpose evaluation is a myth. Although a number of different types of questions can be considered within the bounds of a single evaluation study, it takes meticulous planning and design. Inevitably not even the best-planned study will provide information on all the questions for which people seek answers. Appropriate utilization, interpretation, and dissemination of research findings must be chosen as carefully as all of the other facets of evaluation research.

A number of constraints frequently limit the use of evaluation results. Weiss (1972) suggests at least five which she lists as possible

problems. They are: (a) the evaluator's perception of his role in the utilization process, (b) the organizations' resistances to change, (c) inadequate dissemination of results, (d) the gap between evaluation findings and clear courses for future action, and (e) the tendency of much evaluation to show little or no positive effect.

In Measuring the Effectiveness of Basic Municipal Services, (1974) uses for effectiveness measurements are discussed under these headings: (a) program planning and budgeting, (b) management operations, (c) program evaluation analysis, (d) establishing performance targets, (e) providing meaningful employee incentives, (f) performance contracting, (g) improving citizen feedback for government decision making, and (h) justifying wage and salary increases. The common purpose is to give a government and its agencies information on the question, "How are we doing in meeting goals and objectives for services to the citizens?"

Potential users of the data include the chief administrative officer of the local government and his staff, operating agency managers, the local legislature, and ultimately the citizenry. Individual operating agencies are likely to need more detailed breakdowns on most measures--such as data on smaller geographical areas or more details on specific types of problems (Measuring Effectiveness, 1974).

In order to be used, evaluation results must be communicated (Weiss, 1972). This may sound like a platitude, but it is far from common practice. As noted earlier, most evaluation reports wind up as forty mimeographed copies, four hundred pages long, submitted to a program or a funding agency and piled on a shelf. With notable exceptions, relatively few such reports are transmuted into articles in professional journals or books. Sometimes an agency is responsible, demanding confidentiality,

or the evaluator may simply fail to complete the process of analyzing his results.

As a case in point (Frieden & Kaplan, 1975, p. 169), the Model Cities program distributed large amounts of resources to a limited number of cities. It could be justified only if it generated programs that could be repeated in other cities and lead to future legislation directed at improving the quality of life for all Americans. If the program was to lead to these wider benefits, one would suppose that a sustained and continuous evaluation would be required of its impact.

Surprisingly, evaluation was not given major emphasis in the final task force report. References to it were brief, often oblique. While everyone seemed to assume that evaluation would be part of the program little more than passing reference was made to it during the Congressional hearings. Secretary Weaver implied on several occasions that he had given some thought to measuring the effect of the program, yet he was quite unspecific about the form, content, timing, and methodology of the evaluation. For example, Weaver, before the Congress, noted the need for a purposeful selection of cities and the wisdom of a wide sample. Yet he said nothing about anticipated evaluation (Frieden and Kaplan, 1975).

In an article entitled, "HUD's Attempt at Regional In-house Evaluation," Agelasto (1975, p. 59) portrays a dismal picture. He pictures every facet of the effort as one of such confusion that for every aspect declared positive, the opposite view was also given. Agelasto's most significant observation was, in my opinion, that "regional offices are still run according to the needs and wishes of the regional administrator. . .[who] is preoccupied with performing the management-by-objectives function required of all federal agencies by the Office of Management and Budget."

Perhaps in response to the demands of citizens for accountability, the Committee for Economic Development (CED) indicated in 1971 that: "Without procedures for observing and analyzing the actual performance of programs, the federal government is like a navigator who relies entirely on dead reckoning for a long voyage." In a more recent study of social programs by the Urban Institute, Wholey and his associates (1973) concluded "that substantial work in this field has been almost nonexistent. . . Many small studies around the country have been carried out with such lack of uniformity of design and objective that the results rarely are comparable or responsive to the questions facing policy makers."

The General Accounting Office, in reviewing the performance of economic opportunity programs as directed by the 1967 amendments to the Economic Opportunity Act of 1964, summarized the major requirements for adequate evaluation of those programs as follows:

1. There must be a comprehensive evaluation plan.
2. Evaluation must extend to research and demonstration projects.
3. Evaluation should extend to alternative programs.
4. There must be an adequate evaluation staff.
5. Continuing research must be carried on.
6. Reliable and pertinent data must be available

(CED, 1971).

The CED recommends that Congress authorize funds for evaluation of grant-in-aid programs even in cases where the federal government will not take part in the evaluation directly; in such cases, it should stipulate that the evaluations be made publicly available by the state or local units conducting them.

Why are there research utilization problems? Wholey and his associates (Wholey, Scanlon, Duffy, Fukumots, & Vogt, 1973, p. 50) have noted that the recent literature is unanimous in announcing the general failure to affect decision-making in a significant way. They cite four basic reasons:

1. Organizational inertia. Organizations tend to resist change whereas evaluation usually implies change.
2. Methodological weakness. Policy-makers properly distrust the results of poorly done studies and rely instead on their own experiences or instincts.
3. Design irrelevance. Too many studies bear little or no relationship to the critical program and policy issues.
4. Lack of dissemination. The relevant decision-makers are not shown or briefed on the results of the useful studies.

Buchanan and Wholey (1974) ask several questions. Why is it that almost everyone agrees that evaluation information is essential for good program operation and effective program modification or development and yet evaluation information is only sporadically and inconsistently used for these purposes? Why is it so difficult for evaluators to produce information that can or will be used to develop, improve, or operate social programs?

The response of some evaluators to these questions has been that decision makers simply do not wish to learn the truth about their programs, do not choose to act on the results for political or bureaucratic reasons, or do not know how to use the results when they are available.

The view of decision makers on the other hand, has been that evaluators do not produce information of a sort, or in a form, or at a time that is useful to them in their tasks. Other critics of current evaluation practices believe that evaluation methodology is the problem and that methods must be improved if evaluation information is to be used in policy making.

The evaluator has another problem--bureaucratic survival (Scioli, 1979). The bureaucratic structure is anathema to program evaluation. Once an agency has been set in motion, its first goal is survival--service delivery is secondary. As Scioli writes, "The marriage of the staff to the program in the sacred church of the bureaucracy is a strong impediment to program evaluation."

Dye (1972, pp. 296-297) suggests that there are many reasons why governments do not want to know. Some of the reasons he lists are:

1. Governments pursue incompatible goals and policies at the same time. Overall policy planning or evaluation would reveal the inconsistencies.
2. Many programs and policies have primarily symbolic value. They don't change conditions, but they make people feel that the government cares.
3. Government agencies have a strong vested interest in "proving" that their programs have positive impact. They view evaluation as a threat.
4. Government agencies have "sunk costs"--organizational, financial, etc., and don't want to find out that a given policy does not work.

5. Serious study would involve interference with on-going program activities. Therefore, pressures of every day business take priority.
6. There is wide disagreement about the purposes of programs among administrators, legislators, and clients.

Dye (1972) points out that government administrators and program supporters are ingenious in devising reasons why negative findings about policy impact should be rejected; their "reasons" can be unending. Even in the face of clear evidence that their favorite programs are useless or even counter productive, they will argue that they be maintained.

Pressman (1978, pp. 198-205) offered remedial suggestions at a conference on evaluation research in the following terms:

1. Evaluators should work more quickly.
2. Government should be more interested in long-term problems.
3. Both government officials and outside evaluators should try to identify policy problems before they become crises.
4. Communication and personnel exchange between the user and evaluator communities should be increased.
5. Users should create more incentives for doing high-quality evaluations.
6. Universities should provide greater rewards for interdisciplinary and problem-oriented research.
7. More attention should be paid to the presentation of research in clear and readable form.

Freeman (1974) feels that studies of the utilization and impact of programs in the human resources field may be looked upon in the following ways:

1. As political devices, whether to promote support for an advocate's program or to reduce enthusiasm for an opponent's proposals.
2. As a set of management tools that allow administrators to allocate resources in a systematic and presumably objective way.
3. As public service ventures that permit academicians to delve into the arena of social action in a manner that uses their technical acumen for the social good.
4. As a means of building the knowledge base of that social sciences, for few large-scale experimental studies can be acrualized (and supported) on the basis of scientific worth alone.

In yet another concept of research utilization, Bunker (1978) believes that it must be thought of as any adjustment in the following: (a) observations of the operational setting, (b) definitions of problems or situations, (c) diagnostic interpretations, (d) causal interferences, (e) program proposals, or (f) positions adopted which take into account science findings, concepts, theoretical propositions, or ways of thinking. It is difficult to measure such diverse and diffuse phenomena, he acknowledges, yet it is misleading to treat them as ephemeral or meaningless. Social science knowledge must be tapped to provide both particular substantive guidance and a critical perspective for the latter.

In yet another view, Lippitt (1969, pp. 146-155) perceives the research utilization function as linking agents at various points in the flow of research utilization. New skills of retrieving and organizing research-based knowledge need to link up to the needs of the social practitioner or client population. Helping the practitioner to clarify his resource needs and then helping work through the implications of new knowledge is another aspect of this linking responsibility.

Lippitt (1969) believes that another function of the research utilization agent is to serve as inquiry consultant or trainer, assisting the client population in carrying through its own diagnostic research and working through the meaning of the findings for changes of practice. Also needed are more effective and appropriate ways of linking creative innovations to facilitate the spread and successful adaptation of new practices.

Many program administrators and planners have not had adequate training in the administrative uses of research (Epstein & Tripodi, 1977). Despite this lack, they face increasing pressure from potential funding sources, professional groups and recipients of service to provide and to make use of systematic research data. Moreover, sound professional practice requires that program administrators be more rational and objective in planning, monitoring and evaluating their programs. Research facilitates rationality; it provides information that is essential for responsible administrative decision making. From a pragmatic point of view, research has become an important element in funding proposals because of reduced funding opportunities and increased competition among social programs.

The training of research utilization agents requires a grounding both in behavioral science disciplines, in professional values, and in technology (Lippitt, 1969). This puts a new strain on the fairly segregated curriculum designs and training sequences which still exist in most of our graduate programs.

This leads us to one of the most important processes of science utilization--the use of the scientist as a consultant (Likert & Lippitt, 1953, p. 586). In a vast majority of cases, the effective carrying through of the process of utilization of research findings into integrated policy-making, planning, and operations requires active face-to-face interaction between a social scientist who serves as an interpreter and consultant and the key operating people involved.

Effective communication must be established between relevant social scientific resources and the potential users of these resources (Likert & Lippitt, 1953). One help in this direction is the work of the social-science "middleman," the science writer. A good social-science interpreter is able to classify and synthesize research findings so that they are more clearly related to the problems posed by operating persons.

The utilization agent role is one of consultation and problem solving. Agents help local agencies define their problems and their knowledge needs, retrieve relevant research through resource systems, and aid the agency to adapt the information to their purposes. Weiss (1972) cautions that it is useless to thrust results upon persons who are uninterested, but when people face a decision, they often search for such relevant information.

As evaluation stresses its conceptual as well as its workaday bases and analyzes significant social components, its acceptance will be

greater. For the time being, Weiss (1972) suggests that other outlets must be sought as well. Information briefs, abstracts of evaluation reports, and selective annotated bibliographies may be set up to apprise appropriate audiences about available reports. Personal contact, demonstrations at conferences, and other methods can make information visible and spur the utilization process.

If the specific findings and implications of an evaluation cannot be readily understood, it is not likely to be used (Hatry, Blair, Fisk, & Kimmel, 1976, pp. 24-25). Even the best research analysis will be ignored or rejected if it appears to be esoteric, sloppy, rambling, or incoherent. Findings must be communicated to interested decision makers in very specific matters in a very specific manner. Whether a presentation is written or oral, public officials usually (or often) lack the time or specialized training to pore over lengthy technical arguments, long tables, computer printouts, or formulas to discover what has been found. Findings must be presented in a comprehensible way--in clear English and in a compact and orderly fashion.

Davis and Salasin (1975, p. 663) feel that if an evaluation of attainments of a program proves positive, and if assessment of the unintended side benefits turns out likewise, the dissemination of this information can serve as an important reinforcer to stabilize the change.

Davis and Salasin further express their optimistic views about utilization and communication of evaluation results in the following:

1. Evaluation provides great hope for ultimate beneficiaries of social programs.
2. To fulfill that hope, effective utilization of evaluation is essential and the very field of evaluation may be dependent upon such payoff.

3. The destinies of evaluation results are at least to some extent predictable and the careers of those results can be enhanced.
4. Evaluators should consider extending their role to change consultation.
5. There is a vast body of information on change and models of conceptualization on which to base one's professional services.
6. Utilization of evaluation may most effectively be achieved through the human approach to organizational change.

Scioli (1979, pp. 41-42) considers that perhaps the most limited resource in evaluation is the trained evaluator himself. When one thinks about the number of programs state and local governments deliver, evaluation of them all would require a veritable army of evaluators. Even if this army could be maintained, where would it come from? It used to be an evaluator could recognize a computer and use a statistic he was in business. Today, an evaluator needs a broad interdisciplinary background and a thorough understanding of social science methodology. He needs to know not only the more familiar social science tools such as survey research methods and interviewing procedures, but also time-series analysis, cost-benefit economics, and experimental design principles. Implicit in each of these skills is an appreciation of complex measurement problems and familiarity with multi-variate statistical methodology. Clearly, the trained evaluator is at a premium.

Abert and Kamrass (1974, p. xxiii) assert that "the day of evaluation seems to have arrived." They believe that it would not be an exaggeration to say that it is perhaps "the rage." In social Experiments and Social Program Evaluation, they note that the Office of Management and Budget and

the General Accounting Office are getting strongly into the act. Also Congress itself is both insisting that agencies produce real evidence on the effective utilization of their programs, and providing money to do the evaluations.

A simple illustration of this cresting trend (Abert & Kamrass, 1974) is the fact that in fiscal year 1969 the Department of Health Education and Welfare programmed approximately \$5 million for evaluation. By fiscal year 1971, this had been dramatically increased to \$30 million. The General Accounting Office's 1972 report set specific standards for governmental audits of program results as well as the more traditional checks on legal compliance and fiscal regularity.

## CHAPTER III

### Case Studies in the Mental Health Field

#### Introduction

It seems apparent that there is, in many cases, a basic consistency within the literature of the rapidly growing discipline of evaluation research--in spite of evident differences in purpose, terminology, and methodology. The steps of selecting goals, objectives, and evaluation criteria, collecting and analyzing data, and disseminating findings, for example, are found within most of the evaluation models. The case studies in the work of Hatry et al (1973), Freeman (1974), Tripodi et al (1978), Weiss (1972), and others are described and evaluated within the terms and concepts of these individual authors. Although the emphasis and/or terminology does differ somewhat, an intricate thread weaves many patterns of similarity within their models of evaluation research.

To portray a broader application of the many possibilities within the realm of evaluation research, I would like to present some "real life" evaluations which were done in the field of mental health care in the United States. Under the auspices of the Department of Health, Education, and Welfare, members of the National Institute of Mental Health (NIMH) have developed "Evaluation in practice: A sourcebook of program evaluation studies from mental health care systems in the United States" (Landsberg, Neigher, Hammer, Windle, & Woy, 1979). The authors believe that there has been an overemphasis on research design and methodology and underemphasis on program utility. Their approach is

"aimed at generating timely, practical, and helpful data to allow local agencies to improve their programs." The case studies which follow are extracted from this publication, giving the name of the study and the author or coauthors. Some of the more technical and statistical material is not included for the sake of brevity, but an attempt is made to clearly describe the issues and results.

These studies are not the products of professional "evaluation researchers." They represent an attempt to collect and delineate examples of a broad range of actual mental health program evaluations as practiced under program conditions, rather than select conceptual models or richly funded research or method development studies which most mental health agencies would not be able to replicate.

A Methodology for Allocating Mental Health Equitably .

(Lee DeCola & Willard Van Horne)

CASE I

Goals and Objectives. The purpose of this study was to examine a technical procedure and decisionmaking process developed by the Massachusetts Department of Mental Health for equitably allocating "new monies." There was a need to meet certain requirements of federal enabling legislation, and to determine how equitably or inequitably mental health resources were actually distributed among the 40 catchment areas of the state.

Methodology. The procedure used defined and measured resources, decided how to count non-Department money in tallying an area's funds, and estimated each area's potential "demand" for mental health services.

This approach assumed that estimates of potential demand for mental health services must reflect actual or existing utilization patterns in given areas.

They assumed that (a) socioeconomic factors may represent need but that they are not in themselves measures of need, (b) utilization is an expression of demand for services and therefore a sign of need, but the utilization may also be a response to the presence of professional staff, nearby and effective facilities, or good advertising, (c) it is possible quantitatively to develop a need index based solely on socioeconomic indicators, even when those indicators may have an extremely complex connection with needs for services, and (d) data representing these factors were good enough not only to represent the phenomena referred to, but also to be used for the guiding of policy.

Results. The object was to develop a need index that reflected potential utilization and was based on socioeconomic indicators. The investigators systematically explored relationships among the variables to identify those which might best serve the purpose of developing a need formula. By regressing the admission rate to state hospitals of each catchment area on a variable representing either presence or absence of a community mental health center and five key socioeconomic variables, they were able to account for roughly 56% of the variance in the department variable.

A modified version of the resulting prediction equation was used to estimate predicted inpatient utilization, holding constant the crude proxy measure of service and ability. Each area's predicted demand for mental health services was then simply divided by the total statewide predicted demand, to represent each area's relative share of the total statewide

demand. It was found that 24 of the 40 areas had a resource deficit, while 16 had a resource excess.

A compromise strategy was devised to use the limited financial capacity by giving each area a set amount while distributing a remaining amount among the deficit areas in proportion to their fraction of the total deficit.

Effort. The conduct of the analyses themselves required a modest full-time equivalent personnel investment of about 3 person-months. The manpower effort involved staff activity not only in selection and testing of alternative regression models, but in the more mundane activities of data base updating and the like. A certain amount of additional staff time was expended in surfacing a number of policy-sensitive, definitional issues among top decisionmakers within the department. The actual costs of data preparation, key-punching, and computer processing time was minimal and did not amount to more than \$500.

Impact. To date, the scope of the impact of this equity analysis procedure has been confined primarily to the Division of Mental Health Services. However, the concept of equity is increasingly one which seems to have been firmly anchored in the working vocabulary of central office and regional administrators alike. At a minimum, there appears to be a ready willingness to acknowledge that conventional reliance upon per capita resource allocation strategies are simply inadequate since they do not take into account the differentials in demand attributable to differences in each area's sociodemographic character and resources. In addition, the top decisionmakers in this division have demonstrated strong support for a methodology which apparently "objectifies" an otherwise extremely sensitive political issue.

Implications. The project resulted in two major findings. First, it demonstrated that it is possible for a central office technical planning staff to provide legitimate and acceptable guidelines for the regional allocation of program funds, while preserving local autonomy in the programming and expenditure of these resources. Second, it indicated that even such reasonably sophisticated techniques as computer analysis of data and the use of statistical models can, with sensitivity, be integrated into the decisionmaking system of a mental health bureaucracy.

Suggestions and Recommendations. One of the critical but unavoidable shortcomings of the analysis described was its reliance upon measures of inpatient utilization as a gross measure of overall demand. This was due to the department's management information capacity at the time the analysis was originally conducted. Subsequent demand modeling efforts should use demand estimates which are separately conducted for both inpatient and community-based services. The Department of Mental Health's current management information data now permits more sophisticated treatment of this issue.

A fully developed demand model should take into account private sector capacity of services, and analytical work must be conducted to provide some indication as to which specific services best optimize certain types of policy goals. Only this type of information will provide more satisfactory guidelines to regional and area administrators in allocating resources in a fine-grained way to specific services.

The current operating assumption has been that all essential mental health services are equally important and must therefore be comprehensively developed. Such assumptions, however, have never been rigorously tested, and they ought to be.

## Evaluating Staff and Client Perceptions

(Louis Dwarshuis & Marilyn S. Kolton)

### CASE 2

Goals and Objectives. The purpose of this study was to identify differences in perceptions of interpersonal behaviors between Hispanic counselors and their Hispanic clients in a drug treatment program. In the Hispanic culture, interpersonal relationships are considered more important than in the Anglo culture. One objective was to determine the likelihood that the counselor's perceptions and client's perceptions might influence how the client responds to treatment. Difference in perception by sex is also potentially important, since sex roles are highlighted in the Hispanic culture.

The second objective was to investigate cross-perceptions between female and male counselors and between female and male clients.

A third purpose was to study the interpersonal perception of paraprofessional staff and clients from the same ethnic background. It is often assumed that paraprofessionals are better able to form close working relationships with their clients than professional staff are. There is also an assumption that the client group will be better able to identify with the paraprofessional counselor. The research was designed to test these assumptions.

Methodology. The research instrument, the Interpersonal Check List (ICL), used in this study measures behavior on eight dimensions. Both staff and clients were asked to rate their real self and ideal self. The counselors were also asked to rate the typical client. The counselors were asked to rate their counselor. Clients were given the ICL in English or Spanish by a bilingual researcher from outside the program one month after treatment began.

Results. The staff perceive the clients as considerably more distrustful and more self-effacing than the clients see themselves. An even more pronounced discrepancy in perception is that the clients see themselves as more overconventional and hypernormal than the staff rate the clients. The clients see themselves as significantly more positive in their interpersonal behavior than the staff view them.

An analysis of ratings by sex of staff and by amount of education was also made. Statistically significant differences were found between male and female staffs' ratings of clients in leadership and competitive behavior. Male staff saw the clients as significantly higher on both these traits. Staff with some college, or less education, rate clients as higher on competitive and docile behavior than staff with college degrees. Staff ratings of clients before and after a six month period of training indicated no statistically significant changes over time in perception of the client.

The clients saw the staff more positively than the staff viewed themselves. The clients' ratings of the staff were statistically significant in degree on competitive, aggressive and rebellious characteristics. These findings indicate that the staff see themselves as more self-assertive in their interpersonal behavior, while the client sees the staff as more interpersonally oriented toward others.

The clients rated themselves and the staff very similarly. This finding, however, was not true for the staff, whose perception of the client is not similar to their own self-assessment.

Effort. This study made use of bilingual research interviewers and required 30 work days for administration of the ICL. The costs of the test instrument were insignificant, and the materials were duplicated

without charge. The cost of computer programming to score and analyze the tests was high. Replicating the study in an agency where bilingualism is not a problem and where there is greater understanding of research would require far less effort.

Impact. The results of the study had an immediate impact on the director and staff of the Hispanic drug treatment center. The results broke down the stereotypic belief that the interpersonal relationship of paraprofessional counselors and clients from the same ethnic background would be positive and would automatically facilitate a therapeutic relationship. It brought to the director's attention the need for change in staff training which was then implemented. The staff became more aware of the need for feedback on their perceptions of the clients and how the clients viewed themselves. The director and board gained a better understanding of the need for utilizing professionals to supervise the counselors and to provide training in interviewing and assessment procedures.

The program also decided to utilize more group-oriented forms of treatment and alternatives to individual counseling, such as community involvement. The purpose of these new types of interaction is to increase the positive perceptivity of the clients, the staff, and the community.

Implications. A major finding of the study is the need to assess client and staff perceptions and not to base programs on untested hypotheses about relationships between paraprofessionals and clients or between staff and persons from the same ethnic background. Program evaluation methodologies designed to assess client change as measured by program staff should take into account the staff's perceptions. Training

and supervision directed specifically at the staff-client relationship can assist paraprofessionals in building on their knowledge of the culture to relate to the client.

The results generated by this methodology are especially appropriate for providing feedback to staff on how their attitudes may be affecting their own and the client's behavior in the treatment process.

Suggestions and Recommendations. The method of comparing staff and client perception of each other appears useful in assessing program needs. The method might be more powerful if staff rated specific clients rather than clients in general. Administering the ICL on a periodic basis to staff and clients might be useful in evaluating change over time in staff-client relations.

Three Basic Strategies for Improving the Cost Effectiveness of Social Services (Brian T. Yates)

CASE 3

Goals and Objectives. One popular strategy for improving the cost effectiveness of social service treatments is the development of new theories and new technologies. This approach can bring improvements, but these are not the only, or necessarily the best, way to improve the cost effectiveness of treatment.

One strategy for improving the cost effectiveness of psychological treatments involves operations research--a highly developed technique that can be applied with maximum effectiveness at minimal cost. The studies described below outline three basic paths to improve cost effectiveness that are suggested by the principles of operations research and illustrated by studies in the treatment of obesity.

Methodology. The first strategy for improving the cost effectiveness of a social service system focuses on assessing the degree to which each component of treatment contributes to treatment outcome. For example, one strategy for improving the cost effectiveness of obesity treatment is investigation of the relationship between changes in different eating behaviors and changes in obesity.

The second strategy involves assessing the cost of each component. This strategy may focus on the monetary or the subjective costs of implementing different treatment components (therapist time, equipment, etc.).

A third strategy for improving the cost effectiveness of obesity treatment, given currently available technologies for obesity reduction, involves the use of cost-effective systems for applying the therapy to appropriate behaviors. Such delivery systems can vary greatly in the amount of resources they consume and in the degree to which they translate obesity reduction principles into therapeutic actions.

Results. In the first study the strength and significance of relationships between changes in eating behaviors and changes in obesity were assessed by Pearson, product-moment correlations. All but one eating behavior changed significantly during treatment. There was enough variability of change in behavior to provide the range needed for high correlations.

The second study assessed the subjective cost and benefit of alternative obesity treatment components. Subjects were asked to provide one measure of subjective cost and one measure of perceived benefit for each of the 36 common components of obesity treatments. These data show that there is quite a range of perceived difficulty, but somewhat less of a range of perceived usefulness.

The third study contrasted the cost effectiveness of an intensely delivered (low therapist/client ratio) program for obesity reduction (program X), and high client/therapist ratio lecture-type program (program Y). A one-way analysis of covariance that treated pre-treatment proportion overweight as the covariate found no significant difference in proportional reduction of obesity produced by program X versus program Y. The cost of program X was fixed at \$95 per client, whereas the average cost for a program Y client was \$35.85.

Effort and Impact. The data involved in the three studies were collected from either obesity-clinic archives or questionnaires. A total of approximately three weeks was necessary for data analysis and about \$40 of computer time. The impact on the treatment programs was nil because the programs studied already had terminated. This typical case of "after the fact" evaluation precluded the use of the information gained so that resource allocations could have been changed. More thorough research on component effectiveness has been initiated by clinic personnel. Hopefully, evaluation will be done in conjunction with the study so that the allocated resources can be altered, if indicated, or other changes made to improve the on-going study.

Implications, Suggestions and Recommendations. The author, Brian T. Yates, indicated that these studies are meant to be illustrative--not definitive. The first study used is too small a sample to make any treatment-change decisions based on its findings. He believes the second study makes it clear that subjective costs and subjective benefits can be measured with some validity and are related to some degree even though subjective benefit/cost ratios may differ for strategies. His third study illustrated the powerful difference in cost effectiveness that can be produced by different delivery systems.

These studies illustrate some basic strategies to improve program cost effectiveness; operations research provides a means of combining these strategies with sophisticated mathematical procedures to further enhance the accuracy of decisions about how to change social service systems so as to optimize the cost effectiveness.

A Followup Program For Suicide Attempters: Evaluation  
of Effectiveness (Thomas C. Welu)

CASE 4

Goals and Objectives. A review of the literature reveals convincing evidence that cases of attempted suicide, identified in the emergency room, are usually "treated and released" and receive little or no followup care. This lack of followup care along with the fact that few persons who attempt suicide initiate treatment on their own within the present institutional structure was one of the key reasons for the author's development of a "special outreach program." This program was designed to initiate contact and treatment with a major emphasis on improving continuity and quality of treatment received.

Methodology. All attempted suicides residing in the catchment area of a local Community Mental Health Center (CMHC), brought to the emergency room of a large University Hospital from March 1, 1971, to April 30, 1972, were subjects of this study. The group consisted of a total of 120 attempted suicides with 63 randomly assigned to the experimental group and 57 randomly assigned to the comparison group.

Descriptive data concerning the suicide attempt population were accumulated along with data illustrating the continuity and quantity of care received. An evaluation of the effectiveness of the "special outreach

program" was made by comparing (according to specific evaluative criteria indicating improvement level) the suicide attempters comprising the experimental group receiving the special program with the suicide attempters making up the comparison group. The latter group received the "normal program" that was routinely given and available to persons who attempted suicide. In this "normal program" only 50% of those who had attempted suicide were ever contacted again, leaving the other 50% with no followup care of any kind.

Each suicide attempter was contacted four months after being seen in the emergency room. The basis for comparison was the incidence and prevalence of self-destructive behavior, namely, attempted suicide, excessive use of alcohol, drug misuse, and purposive accidents. Therapists providing the special outreach program were selected from the CMHC. The suicide attempter receiving the special outreach program was contacted as soon as possible after his or her release from the hospital. The exact treatment approach depended upon the needs of the patient and the services available.

It is important to note here that the independent variable in this study was followup, with an emphasis on quantity and continuity of care and not various treatment modalities. The special team member was in weekly, or at least, biweekly contact with the patient throughout the 4 month followup period, either providing the treatment indicated or monitoring the treatment received elsewhere.

Results. Followup and continuity of care given by the normal program were noticeably deficient, as illustrated by the fact that 50% of the comparison research subjects were never contacted after their suicide attempt. The special followup program for the experimental group was very effective in contacting them as well as delivering some type of therapy.

One of the most significant findings resulting from this study was a statistically significant reduction in attempted suicide reattempts by the experimental group when compared to the comparison group at the end of the four month followup period. Out of a total of 62 (no information was available on one person) experimental subjects three reattempted suicide while none of 57 comparison subjects reattempted suicide. The above difference is statistically significant at the .05 level when applying the Fisher's Exact Test.

Thus one is able to say that systematic followup and continuity of care are effective in significantly reducing suicide reattempts in a person who has previously attempted suicide.

Excessive use of alcohol by the experimental group was significantly reduced using the comparison group as a control and the Chi Square test for statistical measurement. Because of the small number of persons classified as "misusing drugs" in the study population, it was not possible to obtain a comparable assessment of treatment effects. Purposive accidents occurred at a relatively equal rate among both groups.

Effort. The team providing the special outreach program consisted of four registered nurses, three social workers (BA & MSW), and two community workers with a high school education. The team members were at the time of the study on the CMHC staff; the monies to support this 2 year study came from a National Institute of Mental Health Grant. Besides the team members listed above, the administrative staff included a principal investigator (75% time) and 3 research assistants (100%, 60%, and 10% time respectively). Total monies spent for the 2 year study were \$66,366.

Much time and effort were spent in obtaining the cooperation of emergency room staff and CMHC staff. Getting to know all key persons in both of the above mentioned areas was essential to the study.

Impact. The impact of this evaluation study on the emergency room staff was indicated by an increased awareness of the needs of suicidal persons. This increased awareness was not measured in any way, however, and the CMHC did not adopt the "special" followup program at the termination of the study. The "normal" treating system promoted by the psychiatric staff continued to exist despite evidence that it was functioning inadequately in the emergency room.

Implications. The primary significance of this research study is the documented evidence empirically demonstrating that a specific program for suicide attempters emphasizing followup and continuity of care does indeed prevent a significant number of suicide reattempts and other self-destructive behavior. One of the more significant results of this study is the emergency of an operational method for the evaluation of the effectiveness of programs whose focus would be the treatment of those who have attempted suicide.

An effective campaign to reduce self-destructive behavior can be achieved by a restructuring of priorities. It should be noted that the positive effect of the special outreach program was achieved by a redirection of energies using the present resources rather than creating a new delivery system of care giving. The practical ramification of this approach is that such a redirection of focus onto the suicide-attempt population can be accomplished with a minimum amount of new staff and expenses.

Suggestions and Recommendations. In order to insure that the special program therapist maintained at least biweekly contact with the research subject in the experimental group, it was necessary for the researchers to contact the therapist every two weeks, asking how the patient was doing and if contact was made. This action was necessary because of the passive and often active resistance on the therapists' part to sometimes "postpone" immediate and/or followup contact with the suicide attempter. To have better control of this resistance it is suggested that members of the special followup team be recruited specifically for the task and not just be given the assignment.

Unfortunately, aggressive seeking out of patients in their homes or elsewhere is not promoted by the providers of mental health services. The mental health care system must be prodded to forego the easily accessible patients in favor of the "hard-to-reach" persons who have made suicide attempts.

## CHAPTER IV

### Comparative Cases and Comments

#### Analysis of Case 1 (A Methodology for Allocating Mental Health Resources Equitably)

The Massachusetts Department of Mental Health (DMH) conducted this research study. Their concern was to develop a more equitable distribution of monies for mental health services within the state's regions and areas. The DMH developed a formula for distribution by adding together inpatient use, socioeconomic indicators, and service availability in an area. This sum was divided by statewide "predicted demand." Only dollars, areas, and "statistical demand" were utilized to determine equity in the mental health field.

The last paragraph of the case study contains the amazing conclusion that "all mental health services are equally important" and therefore must be equally developed--the results of their own research notwithstanding. It is admitted that this finding was not tested by the research.

In my opinion, the study ended at the point where it should have started, i.e. evaluating the needs perceived by individual clients and those needs which the community considered to be of importance. The study concludes with the recommendation that further research be done in analytical terms. Perhaps perceived needs are implied in their planned additional analytical work "to indicate which specific services best optimize certain types of policy goals." No federal bureaucrat could say less in finer rhetoric.

It is strange that a research effort to allocate mental health resources equitably became instead a statistical exercise. The perceptions by individuals and/or communities about any existing mental health programs or specific problems or needs were ignored. Playing the numbers game may appear to make the decisions more objective, and a given amount of money was stipulated to be divided among the more needy areas. However, many states have areas of severe poverty which greatly surpass the number of people in comfortable circumstances. There should be a followup study to ascertain the levels of need for services in each area in proportion to the numbers of people and resources in each area.

#### Analysis of Case 2 (Evaluating Staff and Client Perceptions)

The setting for this research was a drug treatment center. Both the clients and the paraprofessional staff, who work directly with the clients as counselors, were Hispanic. This precluded any problem factor due to differences in ethnic background, but the research revealed that previous assumptions about automatic positive relationships due to shared ethnic origins were invalid.

The stated purpose of this research project was to identify and evaluate differences in perception because of roles, education, and sex. Again the research revealed that previous assumptions relating to roles, education, and sex were also invalid. A basic conclusion was reached that decisions should not be made on the basis of untested "assumptions."

An important conclusion reached from this study was that the staff and the clients' self-ratings and perceptions of each other were very different. This finding made an "immediate impact" on the director and staff, and adjustments were made in training and programming. This is such a refreshing sentence to write! The insight needed to test for

differences of perceptions and the willingness to make immediate changes to improve the program are rather rare. Few people are willing to acknowledge their own fallibility.

Unlike Case 1, the concern was for immediate and beneficial results for the individuals of both staff and clients. The changes were made primarily to enhance the treatment of the clients. However, since one of the changes was to expand group-oriented programs by planning for greater community involvement, the perceptions of the drug treatment program by the community were also changed.

Analysis of Case 3 (Three Basic Strategies for Improving the Cost Effectiveness of Social Services)

This case is a straight forward study of methods for improving the cost effectiveness of a social service program which involved obesity treatments. Three methods are described. The research strategies involve (a) measuring the effectiveness of each major treatment component, (b) assessing the objective and subjective costs of implementing each treatment component, and (c) contrasting the cost effectiveness of the treatment delivery system as well as the cost effectiveness of the treatment technology.

A major portion of the documentation for this research is written in the vernacular of comparative statistics, a language with which I am only vaguely familiar. Details of the methodology involved, statistical formulas utilized, and cost effectiveness ratios are indicative of great sophistication in the researcher. It is precisely written and informative for a lay person, but would be more valuable to another researcher in the field of cost analysis.

This is a good example of the accumulated wisdom which can and should be used as background for further research. It illustrates some basic strategies to improve program cost effectiveness by combining operations research and mathematical procedures to aid in making decisions involving cost effectiveness. Both objective and subjective costs are considered which properly balances social and fiscal responsibilities.

"Cost effectiveness" is not a term which projects the usual social and psychological concerns such as the equitable allocation of resources in Case I or the perceptions of clients and staff in Case 2. Nevertheless, the opportunity for quality evaluation research depends upon how effectively monies are utilized--and this means a constant search for cost effectiveness in every research project.

Analysis of Case 4 (A Followup Program for Suicide Attempters:

Evaluation of Effectiveness)

This four month followup outreach program was developed with an emphasis on continuity and quantity of treatment given to suicide attempters after being seen in a hospital emergency room. Persons who have attempted suicide are a high-risk group in relation to ultimately completing suicide, but they are usually "treated and released" with little or no followup care. Suicide attempters from a specified time frame were randomly assigned to the "special outreach program" or the "normal" treatment program. This gave the study validity when the evaluation was made by comparing the two groups at the end of a four month period using specific evaluative criteria.

One of the most significant findings was a statistically significant reduction in suicide reattempts by the experimental group when compared to

the comparison group. With full knowledge of this fact, the "normal" treating, which provided for little or no followup care, was continued by the hospital's psychiatric staff.

Followup and continuity of care were supposed to be a part of the "normal program" but it was quite deficient in practice. Only 50% of the comparison group were contacted while all but one of the experimental subjects were contacted and followed. Neither the staff nor others involved in this mental health care system appeared to be interested in helping suicidal persons. Few people, including the medically trained personnel, have any concept of their pain and feel no empathy for them. The void in understanding and acceptability between "mental illness" and physical illness is still great, even in the medical community.

Cases 1, 2, and 3 covered the topics of the equitable allocation of monies for mental health, the perceptions between staff and clients, and a study of cost effectiveness in social service programs. All of them were presented as viable concepts, ready to take roots and grow if not completely established previously. It appears tragic to me that, in the face of documented evidence empirically demonstrating that a specific program does prevent suicide reattempts, no effort to utilize or even retain the program was made. It makes me ask what the concept of social service really means in our culture. Can we only provide service in those situations which allow us to feel comfortable?

#### Comments

With the recognition that governmental responsibility for social programs is here to stay, despite a substantial protest from many individuals and groups who view many expensive programs as nonproductive, the demand for valid research evaluation continues to escalate.

Although there are substantive differences over the proper approach to evaluation research, there are many similarities involved in the use of varied social indicators. This leads many social researchers, organizations (public and private), and citizens to anticipate the formulation of well defined social indicators in order to compare the effects of a program with its stated goals. Economic indicators have long been an acceptable tool for evaluating and projecting trends in the economic world. There is an enormous need for valid, commonly accepted social indicators to more adequately measure present social programs and conditions. These indicators could also serve to implement decision making for future programming.

It is the responsibility of the researcher, in concert with the appropriate parties, to select an evaluation design that can be used to provide relatively unbiased information pertinent to the questions of evaluation. In my opinion, evaluation strategies would produce more accurate guidelines if there were inputs from more of the parties involved--researchers, data processors, decision makers, as well as the target population (individual or community). This would be difficult to accomplish, perhaps impossible, but if the viewpoints of more sectors could in some way be projected into the planning of an evaluation research program, many facets of the research findings would be enhanced in a very meaningful way.

In Case 2 there was an attempt to measure the different perceptions of the staff and clients. A clear indication was found that the perceptions of the participants were quite different from the assumptions on which the drug treatment program had been built. The variations were documented by administering the Interpersonal Check List which measured behavior in eight dimensions.

There are many psychological tests for measuring attitudes, needs, goal setting tendencies, etc. A multitude of tests are available to measure to some extent almost any area of social concern. An acknowledgment of the wide spectrum of perceptions between the researchers, the individual client, and an entire population group would reveal a vastly different array of needs, goals and other priorities.

In most cases there are at least four major viewpoints or perspectives among the participants of an evaluation research project. These include (a) the sponsor which funds or initiates the program, (b) the provider or agency which delivers the services, (c) the consumer or client who utilizes the program or service, and (d) the community or residents of the service area. In comparing or contrasting these four participants in an evaluation project, a focus on multiple and differing perspectives of the optimum in utilization of time, money, and other resources could add a dimension of great importance to the impact on social action and human service programs. To have this desired result requires the collective wisdom and cooperation of all four groups.

Adding these perspectives, plus a followup component, to a research program when it is initially being structured could divulge more applicable methodologies which would in turn produce more accurate findings. This could redirect or totally change the thrust of the research; the findings also might more clearly delineate the heart of a problem and reflect new directions and needs for future research. A generalized matrix incorporating the perspectives of these four parties could be depicted in the following form:

COMPONENTS	SPONSOR	PROVIDER	CONSUMER	COMMUNITY
GOALS				
OBJECTIVES				
CRITERIA				
INDICATORS				
RESEARCH DESIGN				
DATA COLLECTION				
DATA ANALYSIS				
EFFORT				
IMPACT				
FOLLOWUP				

There is a world of difference between identifying facts and identifying with feelings. Both are necessary for true comprehension. Facts must be accurate and are of the utmost importance, but the perceptions of people are also real and must be taken into account if social research is to be meaningful and successful. Perspectives are facts, in essence, to an individual at a particular moment of time and life-space. To ignore these perspectives in an evaluation effort is to push aside an important dimension of social research.

Another facet of evaluation research which has been ignored by most researchers is measured differences which occur because of a time factor. What are the short and/or long range effects of a social program? Who will be affected by a program now, having been involved in a similar or greatly different program in the past? Can future ramifications be projected so that an ameliorating feature can be built into a new program? If there is no followup, can it be assumed that the findings of a study will remain the same? Change directions? Stultify previous efforts?

I feel that researchers must attempt to find and describe both the dimensions of facts, perspectives, and changes over time in social service programs. These are prerequisite for fully creating and portraying an acceptable social phenomenon or accurately transcribing our situation of confusion into one of integrity and validity.

## CHAPTER V

### Evaluation Research and Additional Dimensions

#### Evaluation Research as Synthesized from the Literature

Within the majority of the literature pertaining to evaluation research, the basic pattern or procedure usually includes the following steps, described only in brief detail:

1. Determine the broad goals of research program. Identify issues to be resolved, activities to aid management, program planning, etc.
2. Refine the goals into more specific objectives. Decide who is expected to perform which tasks. Define when, where and why particular information is needed.
3. Establish evaluation criteria. Determine whether the program is achieving its objectives in order to be continued, reduced, or eliminated.
4. Search for appropriate social indicators. These are often found in social and demographic statistics in public records. They may be used to measure factors in a time series or changes in various situations.
5. Determine a proper research design. This is possibly the most technical part of evaluation research. Successful design depends on adequate experience and knowledge to take the initiative in selecting and developing plausible and consistent information.

6. Collect and analyze data. Much data, available in data banks and a variety of other places, may be analyzed comparisons and statistical formulas. Other data may be analyzed by trained observation techniques or computer programs.
7. Determine cost effectiveness. Assess the degree of contribution in relation to the amount of cost. Sophisticated systems involve various cost analysis formulas.
8. Prepare results for utilization. Make concise reports in an interesting format and understandable language. Report only relevant material and present it in an organized manner.

#### Additional Dimensions

There are at least two dimensions of evaluation research which are not given sufficient attention for my satisfaction. They are (a) the perspectives of each participant, and (b) planned followup evaluation research over varying periods of time.

If the first dimension were to be carefully considered and added to the beginning of a project, and the second dimension developed at the end of every evaluation program, the findings would become more and more fruitful as the program is implemented. "Facts" often take on widely ranging implications as they are viewed from the many perspectives of status, education, age, sex and many other such factors. While there is still much to be learned from social, psychological, sensitivity, and other tests, our programs could be greatly improved if the knowledge we already have were to be implemented.

It is almost too difficult to consider trying to change anything in our land of endless bureaucracy, but the restless waves of discomfort, anxiety, and anger are becoming widely felt perspectives which are becoming more apparent as social and economic needs are not being met. Our current status, in many ways, resembles the early sixties. Some are disenchanted for those same reasons--from their perspective changes have been made primarily on the surface.

Planned followup evaluation research, from former and current studies, could more graphically delineate problems which need attention. Social research is not conducted for its own sake in a scientific vacuum. It is done to find answers to specific questions in order to overcome unresolved social problems.

The ultimate value of research evaluation depends upon the degree to which it is utilized. Its most common purpose is to give a government and its agencies answers to the questions, "How well are we accomplishing our stated goals for serving our citizens? Should a given service or program be bolstered, changed in emphasis, or deleted?" Plausible answers to these questions could be better determined by routine or specifically planned followup evaluation research performed by highly trained and competent research utilization agents. These professionally trained people are the most scarce resource in evaluation research.

Training for research utilization agents, just as for the research evaluator, must incorporate skills from the social and behavioral sciences, acumen in research techniques, and keenness in diagnostic capacity. Both must work through the meaning of findings and determine the most effective methodologies for utilization. There is increased interest in evaluation in many universities, especially in the political

science departments. Some institutions even have an organized interdisciplinary policy studies program. Hopefully, this will result in more graduates entering government who are knowledgeable about the intricacies of public policy and skilled in facilitating evaluation research.

In many sectors, the realization has come that social programming must be improved. The fact that evaluation results are becoming more valid and are being taken more seriously is only an indication that more fundamental social experimentation and evaluation must be done.

### Conclusions

Carrying out a good evaluation program is a complex and fragile process which can be done to the highest degree only when evaluators, constituent parties, and decisionmakers are fully aware of the difficulties, limitations and efforts involved in such an attempt. An evaluation is most effective when it is visibly related to a particular problem or target population to be studied or served by the evaluation program. Evaluations must be skillfully coordinated with every party involved at every stage of the implementation if the evaluator ever expects the results to be used.

There is no ultimate or ideal evaluation; all have weaknesses. Perhaps the most pragmatic approach for the evaluator is (a) to anticipate key decision issues so that he can provide timely data, and (b) to aim at giving the participants a profile of services by using multiple evaluation techniques and several outcome measures.

Evaluations by themselves never determine what decisions will be made. Decisions involve subjective and normative values which are of necessity independent of objective data to a degree. Perhaps an

investigation of the perceptions ingrained in each of the participants would improve the evaluators ability to define the basic problems more effectively. Evaluation research raises the level of discussion and increases the accuracy of the data used in the process. Every evaluation will yield returns in direct proportion to the quantity and quality of dimensions researched.

As I noted before, evaluation research, in and of itself, has no magic. Evaluation studies can be no better than the abilities and efforts of those persons carrying out the research and utilizing the results.

REFERENCES CITED

- Abert, J. G., & Kamrass, M., (Eds.) Social Experiments and Social Program Evaluation. Cambridge: Ballinger, 1974.
- Agelasto, M. A. II., HUD's Attempt at Regional In-house Evaluation: A Lesson Learned. Evaluation, 1975, 2(2), 55-59.
- Attkisson, C. C., Hargreaves, W. A., & Horowitz, M. J. Evaluation of Human Service Programs. New York: Academic Press, 1978.
- Bauer, R. A. (Ed.) Social Indicators. Cambridge: M.I.T. Press, 1966.
- Bauer, R. A. Social Indicators: Or Working in a Society Which Has Better Social Statistics. In S. Anderson, (Ed.), Planning for Diversity and Choices: Possible Futures and Their Relations to the Man-controlled Environment. Cambridge: M.I.T. Press, 1968.
- Bell, R. A., Nguyen, T. D., Warheit, G. J., & Buhl, J. M. Social Utilization, Social Indicator, and Citizen Survey Approaches to Human Services Need Assessment. In C. Attkisson, & W. A. Hargreaves. Evaluation of Human Services Programs. New York: Academic Press, 1978.
- Bennis, W. G., Benne, D., & Chin, R., (Eds.). The Planning of Change. (2nd Ed.) New York: Holt, Rinehart and Winston, 1969.
- Branch, M. C. Critical Unresolved Problems of Planning Analysis. Journal of American Institute of Planners, 1978, 1, 47-59.
- Brownstein, C. N. The Experimental Evaluation of Public Policy. In C. O. Jones & R. D. Thomas (Eds.), Public Policy Making In A Federal System. Beverly Hills: Sage, 1976.

- Buchanan, G. N., & Wholey, J. S. Overview of Evaluation at the Federal Level. In J. G. Abert & Kamrass (Eds.), Social Experiments and Social Program Evaluation. Cambridge: Ballinger, 1974.
- Bunder, D. R. Organizing To Link Social Science With Public Policy Making. Public Administration Review. 1978, 38(3), 223-231.
- Campbell, D. T., & Stanley, J. Experimental and Quasi-experimental Designs for Research. Chicago: Rand McNally, 1963.
- Committee for Economic Development. Improving Federal Program Performance. New York: Author, 1971.
- Davis, H. R., & Salasin, S. E. The Utilization of Evaluation. In E. L. Struening & M. Guttentag (Eds.), Handbook of Evaluation Research. Beverly Hills: Sage, 1975.
- Deming, W. E. The Logic of Evaluation. In E. L. Struening & M. Guttentag (Eds.), Handbook of Evaluation Research. Beverly Hills: Sage, 1975.
- Dye, T. R. Understanding Public Policy. Englewood Cliffs, New Jersey: Prentice-Hall, 1972.
- Edwards, C. W., & Yarvis, R. M. Let's Quit Stalling and Do Program Evaluation. Community Mental Health Journal, 1977, 13(2), 205-211.
- Epstein, I., & Tripodi, T. Research Techniques for Program Planning, Monitoring, and Evaluation. New York: Columbia University, 1977.
- Etzioni, A., & Lehman, E. W. Some Dangers in "Valid" Social Measurement. In B. M. Gross, Social Intelligence for America's Future: Explorations in Societal Problems. Boston: Allyn and Bacon, 1969.
- Festinger, L. Laboratory Experiments. In L. Gestinger & D. Katz (Eds.), Research Methods in the Behavioral Sciences. New York: Dryden, 1953.

- Fitzpatrick, R. The selection of Measures for Evaluating Programs. In Evaluation Research: Strategies and Methods. Pittsburgh: American Institute for Research, 1970.
- Fox, K. A. Social Indicators and Social Theory: Elements of An Operational System. New York: John Wiley, 1974.
- Freeman, H. E. Evaluation Research: A Growth Industry. In J. G. Abert & M. Kamrass (Eds.), Social Experiments and Social Program Evaluation. Cambridge: Ballinger, 1974.
- Frieden, B. J. & Kaplan, M. The Politics of Neglect: Urban Aid From Model Cities to Revenue Sharing. Cambridge: The M.I.T. Press, 1975.
- Glass, G. V. (Ed.). Evaluation Studies Review Annual. (Vol. 1), Beverly Hills: Sage, 1976.
- Goldenberg, E. N. Evaluating Municipal Services. Public Administration Review, 1979, 39(1), 94-98.
- Guttentag, M., & Saar, S. (Eds.). Evaluation Studies Review Annual, (Vol. 2). Beverly Hills: Sage, 1977.
- Hatry, H., Blair, L., Fisk, D., & Kimmel, W. Program Analysis for State and Local Governments. Washington, D.C.: Urban Institute, 1976.
- Hatry, H. P., Winnie, R. E., & Fisk, D. M. Practical Program Evaluation For State and Local Government Officials. Washington, D.C.: Urban Institute, 1973.
- Havelock, R. G. Research on the Utilization of Knowledge. University of Michigan, 1974.
- Hoole, F. W. Evaluation Research and Development Activities. (Vol. 68). Beverly Hills: Sage, 1978.
- Hoshino, G. Social Services: The Problem of Accountability. In G. Slavin. Social Administration: The Management of the Social Services. New York: Haworth Press & Council on Social Work Education, 1978.

- Janis, I. L., & Mann, L. Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment. New York: Free Press, 1977.
- Johnson, G. H. The Purpose of Evaluation and the Role fo the Evaluator. In Evaluation Research: Strategies and Methods. Pittsburgh: American Institute for Research, 1970.
- Land, K. C., & Felson, M. A Dynamic Macro Social Indicator Model of Changes in Marriage, Family, and Population in the United States: 1947-1974. Social Sciences Research, 1977, 6, 328-362.
- Landsberg, G.; Neigher, W. D., Hammer, R. J., Windle, C., & Woy, J. R. (Eds.). Evaluation in Practice: A Sourcebook of Program Evaluation Studies from Mental Health Care Systems in the United States. (DHEW Publication No. (ADM) 78-763), Washington, D.C.: U.S. Government Printing Office, 1979.
- Lichfield, N., Kettle, P., & Whitbread, M. Evaluation in the Planning Process. New York: Pergamon Press, 1975.
- Likert, R., & Lippitt, R. The Utilization of Social Science. In L. Festinger & D. Katz (Eds.). Research Methods in the Behavioral Sciences. New York: Holt, Rinehart & Winston, 1953.
- Lippitt, R. The Process of Utilization of Social Research to Improve Social Practice. In W. G. Bennis, K. D. Benne, & R. Chin (Eds.). The Planning of Change. (2nd Ed.). New York: Holt, Rinehart & Winston, 1969.
- Lynn, L. E., Jr. Policy Relevant Social Research: What Does It Look Like? In M. Guttentag & S. Saar (Eds.), Evaluation Studies Review Annual, (Vol. 2). Beverly Hills: Sage, 1977.
- Martinson, R. California Research at the Crossroads. Crime and Delinquency. 1975, 22, 180-191.

- Measuring The Effectiveness of Basic Municipal Services: Initial Report.  
Washington, D.C.: The Urban Institute and International City  
Management Association, 1974.
- Nunnally, J. D., & Wilson, W. H. Method and Theory for Developing  
Measures in Evaluation Research. In E. L. Struening & M. Guttentag  
(Eds.), Handbook of Evaluation Research. Beverly Hills: Sage, 1975.
- Prather, J. E., & Gibson, F. K. The Failure of Social Programs. Public  
Administration Review, 1977, 37(5), 556-564.
- Riecken, H. W. Social Experimentation. In C. Abt (Ed.), The Evaluation  
of Social Programs. Beverly Hills: Sage, 1976.
- Riecken, H. W. The Volunteer Work Camps: A Psychological Evaluation.  
Reading, Mass: Addison-Wesley, 1952.
- Rivlin, A. M. Systematic Thinking for Social Action. Washington, D.C.:  
Brookings Institute, 1971.
- Rocheleau, B. A. Without Tears or Bombast: A Guide to Program Evaluation.  
Dekalb, Ill.: Center for Governmental Studies, 1975.
- Rosener, J. B. Citizen Participation: Can We Measure Effectiveness?  
Public Administration Review, 1978, 38(5), 457-463.
- Salmon, L. M. Follow-ups, Letdowns, and Sleepers: The Time Dimension in  
Policy Evaluation. In C. O. Jones & R. D. Thomas (Eds.), Public  
Policy-Making in the Federal System. Beverly Hills: Sage, 1976.
- Scioli, F. P. Problems and Prospects for Policy Evaluation. Public  
Administration Review, 1979, 39(1), 41-45.
- Struening, E. L. & Guttentag, M. (Eds.). Handbook of Evaluation Research.  
Beverly Hills: Sage, 1975.
- Tripodi, T., Fellin, P., & Epstein, I. Differential Social Program  
Evaluation. Itasca, Ill.: F. E. Peacock, 1978.

Warwick, D. P. & Lininger, C. A. The Sample Survey: Theory and Practice.

New York: Human Sciences Press, 1976.

Weiss, C. H. Evaluation Research: Methods for Assessing Program

Effectiveness. Englewood Cliffs, N. J.: Prentice-Hall, 1972.

Wholey, J. The Role of Evaluation and the Evaluator in Improving Public Programs: The Bad News, the Good News, and a Bicentennial Challenge.

Public Administration Review, 1976, 36(6). 679-683.

Wholey, J. S., Scanlon, J. W., Duffy, H. G., Fukumots, J. S., & Vogt, L. M.

Federal Evaluation Policy: Analyzing the Effects of Public Programs.

Washington, D.C.: Urban Institute, 1973.

Wrightsman, L. S. Social Psychology. (2nd Brief Ed.). Monterey, CA.:

Brooks/Cole, 1977.

## BIBLIOGRAPHY

- Abt, C. C. (Ed.). The Evaluation of Social Programs. Beverly Hills: Sage, 1976.
- Blackwell, B. L. & Bolman, W. M. The Principles and Problems of Evaluation. Community Mental Health Journal, 1977, 13(2), 183.
- Boen, J. R. Will Evaluation Outgrow the Game Stage? Evaluation, 1975, 2(2), 9-10.
- Caplan, N., Morrison, A., & Stambaugh, R. J. The Use of Social Science Knowledge in Policy Decisions at the National Level. Ann Arbor: Institute for Social Research, 1975.
- Golembiewshi, R. T., Gibson, F., & Cornog, G. Y. (Eds.). Public Administration: Readings in Institutions, Processes, Behavior, Policy. (3rd Ed.). Chicago: Rand McNally, 1976.
- Havelock, R. G. Planning for Innovation Through Dissemination and Utilization of Knowledge. Ann Arbor: Institute for Social Research, 197.
- Kerlinger, F. N. Foundations of Behavioral Research. (2nd Ed.). New York: Holt, Rinehart & Winston, 1973.
- Lyons, G. M. (Ed.). Social Research and Public Policies: The Dartmouth/OECD Conference. Hanover, H. H.: University Press of New England, 1975.
- Redburn, F. S. On "Human Services Integration," Public Administration Review, 1977, 37(3), 264-269.
- Sheldon, E. B. & Freeman, H. E. Notes on Social Indicators: Promises and Potential. Policy Sciences, 1970, 1, 97-111.

Shonfield, A. & Shaw, S. (Eds.). Social Indicators and Social Policy.

New York: Crane, Russak, 1972.

Simon, J. L. Basic Research Methods in Social Science. New York:

Random House, 1969.

Slavin, S. (Ed.). Social Administration: The Management of the Social  
Services. New York: Haworth Press & Council on Social Work

Education, 1978.