**Working Paper 179**

# The Evaluation Process in State Vocational Rehabilitation Agencies: Comments for the Prime Study Group on Program Evaluation, Tenth Institute on Rehabilitation Services

Frederick Collignon
Barbara Thompson

June 1972

University of California at Berkeley
**$5.00**

**Working Paper 179/ R3007**

# The Evaluation Process in State Vocational Rehabilitation Agencies: Comments for the Prime Study Group on Program Evaluation, Tenth Institute on Rehabilitation Services

Frederick Collignon
Barbara Thompson

*The research reported here is being supported by a grant from the Rehabilitation Services Administratin of the Social and Rehabilitation Service, U.S. Department of Health, Education and Welfare.*

June 1972

University of California at Berkeley
The Institute of Urban and Regional Development

# PREFACE

This working paper consists of a series of comments which were prepared for consideration by the members of the Prime Study Group of the Tenth Institute on Rehabilitation Services (I.R.S.), which was mandated "to develop criteria and methodology for evaluating the effectiveness and quality of services within State Vocational Rehabilitation programs." The twelve members of the Study Group initially met in Charleston, West Virginia in November, 1971, to discuss how they would respond to their mandate. At this meeting, a final report was outlined and various members of the Study Group assumed responsibility for submitting first drafts of particular chapters. The chapters in the original outline corresponded to the chapters listed in the Table of Contents, with the exception that the Study Group outline included a seventh chapter on organizational problems of utilizing the findings of program evaluation efforts. Each member of the Study Group was invited to submit ideas to other Group members for their various chapter drafts. The comments presented in this working paper were written by ourselves and submitted by Professor Collignon, a member of the Study Group, to the other Study Group members. In addition to the comments, we submitted a longer paper as a proposed draft for the third chapter on Issues and Criteria in Evaluation. The longer paper is available in the Institute series as Working Paper No. 173/RS003, entitled "Guidelines and Criteria for Evaluating Vocational Rehabilitation Programs."

Although we have engaged in some editing and splicing together of the original comments, we have intentionally not attempted to pull together our approach and recommendations for program evaluation into a polished, synthesized paper addressed to state agencies. We feared that the issuance of such a paper might be perceived by some as competing with the efforts of the Prime Study Group and I.R.S. We believe that state agencies should look to the I.R.S. for such synthesis. We do believe that the comments presented in this working paper will be useful to many state agencies as they internally analyze the I.R.S. document, ponder how to set up evaluation units, and look for supplemental material in training or orienting new program evaluation staff.

The suggestions presented in these comments are those solely of the authors. By no means do the comments represent the judgment or opinions of the Prime Study Group. At the same time, we must acknowledge the insights which we gained from the other members of the Prime Study Group; many of these insights are reflected in these comments. We would particularly acknowledge our indebtedness to two state agency members of the Study Group, Mr. W.F. Morehead, Assistant Director for Program Evaluation, Texas Rehabilitation Commission, and Mr. Adam Zawada, Director of Planning and Research in the State of Florida's Division of Vocational Rehabilitation. Several of our recommendations borrow heavily from their suggestions in our I.R.S. discussions. Their anecdotes and insights gained from extensive experience in program evaluation activity in two important state agencies stimulated our thinking and understanding of the problems of program evaluation in state agencies.

We would also acknowledge the helpful criticisms of Dr. Michael Teitz and the postdoctoral and graduate students who reviewed these comments and discussed them at length with the authors in the Workshop on

Program Evaluation and Policy Analysis of rehabilitation Services sponsored by the Department of City and Regional Planning, University of California at Berkeley.

Background of the Prime Study Group Effort and I.R.S.

Both the Institute on Rehabilitation Services and the particular Prime Study Group for which this paper was written are rather unique institutions. The Institute was initially developed through the joint planning of the Council of State Administrators of Vocational Rehabilitation, the Vocational Rehabilitation Administration, and leading members of the former Guidance, Training, and Placement Workshops. Objectives of the I.R.S. include:

1. Identification of problem areas in the rehabilitation process.
2. Development of methods for resolving identified problems.
3. Development of methods for incorporating solutions into state programs.

Each year, an I.R.S. Planning Committee, composed of state agency administrators, leaders of rehabilitation professional associations, and Department of Health, Education and Welfare (D/HEW) officials, meet in Washington to identify several major problems in the rehabilitation program requiring attention. One or more Prime Study Groups, consisting of about a dozen members each, are then designated to study and prepare a report on each problem identified. Members of the Prime Study Group are drawn primarily from state agencies. Several Social and Rehabilitation Services (S.R.S.) Regional Office and Rehabilitation Services Administration (R.S.A.) staff actively work with the Prime Study Groups. A university sponsor is also designated for each Study Group to provide technical and writing assistance. The drafts which emerge from the Study

Group thus incorporate the best thinking of individuals actively working within rehabilitation programs. The drafts are tempered by the experience and wisdom of agency personnel who have spent years in the program. Indeed, many different perspectives are brought into each Study Group: Federal, state central office, field office and university. The drafts prepared by the Study Groups are then discussed in detail at a workshop attended by state agency staff from across the country. The suggestions made at these workshops are incorporated by the Study Group into its draft, the draft is rewritten, and a final presentation is made to the National Rehabilitation Association (N.R.A.) annual convention. When the reports are subsequently published, they are used as training materials in university and agency training programs for counselors and administrators. Because the materials have been prepared and reviewed at length by individuals drawn from all levels within the rehabilitation system, the materials are perceived as having come out of the program and the state agencies themselves rather than being developed by academic researchers or imposed by Federal agencies. This process by which the materials were developed bestows authority and legitimacy upon the materials. This process is, we believe, almost unique within social service and social action programs.

The Prime Study Group on program evaluation is unique because of the subject upon which it focuses and because of the anticipated use of the final report. Evaluation was identified by the I.R.S. Planning Group in the summer of 1971 as one of the key challenges confronting rehabilitation agencies. The problem, however, was not one which readily could be translated into a need for training materials, the traditional orientation of I.R.S. study groups. The Planning Committee decided

nevertheless to sponsor a Study Group, a group which represented the first effort of state rehabilitation agencies to address collectively the needs and problems of program evaluation. The anticipation of the Prime Study Group is that their report will be used by State Agency Directors and evaluation staff as suggested basic guidelines for organizing and operating an evaluation unit. Because the pending renewed Vocational Rehabilitation legislation stresses the need for program evaluation and because Federal money is increasingly being made available to state agencies to create evaluation units and expand their evaluation activities, the guidelines suggested by the Study Group should provide significant assistance to state agencies.

Because program evaluation is a priority concern of state rehabilitation agencies, the work of the Prime Study Group has aroused great interest. The report will be presented to the annual convention of the National Rehabilitation Association in Puerto Rico during the summer of 1972 and will subsequently be published in 1973 by the Department of Health, Education, and Welfare and made available to state agencies and universities. The document will cover the reasons for evaluation, the structure and sequential steps of the evaluation process, an overview of the issues and criteria in evaluating vocational rehabilitation programs, suggestions for organizing an evaluation unit (e.g. staff needs, relationships with other agency units, procedures for launching evaluation studies), methodological suggestions, and guidelines facilitating the implementation of evaluation findings. All interested readers of the comments in this working paper should examine the final I.R.S. report when it emerges.

TABLE OF CONTENTS

Chapter IV. (cont.) Page

# I. WHY DO EVALUATION?

Agencies and researchers may use flowery words in describing why evaluation is needed or sponsored, but such words often conceal the actual reluctance and fear of agencies in consenting to, much less initiating, evaluation. Often program evaluations are sponsored because a legislature, higher executive office, or Federal funding source has <u>required</u> evaluation. These involuntary evaluations, when conducted by the agency being evaluated, become perfunctory. A report is generated, furnished to the outside agency requiring the evaluation, and then relegated to the shelf. As with many requirements that applications for funding show evidence of careful planning, the paperwork which is the basis for demonstrating compliance with the requirements has no real impact on the program.

When agencies voluntarily decide to sponsor an evaluation, the agency is usually prompted by a desire for help in one or both of two key organizational functions: making some decision amidst uncertainty, and legitimating changes. Let's examine these two functions more closely.

1. To decide among alternatives and to guide future actions. When choices are to be made, as in deciding among alternative projects and strategies, or in setting priorities, information on the success and value of past activity can be useful. Evaluation is only one input to the planning and decision-making process, however. It, nevertheless, should be a key input. Agencies usually spend resources on evaluation for decision-making only when there is a fair amount of uncertainty about what decision ought to be made.

2. To legitimate changes in practice and program to:

   a. People within the rehabilitation system -- if change is to take place within an organization, then the personnel who make up that organization will need to understand why change is desired. This is a first step toward the process of eliciting the acceptance and understanding of change on the part of those who will be expected to implement changes.

   b. People outside the rehabilitation system -- changes and improvement will also require the support of legislators, governors, the general public, and other outsiders.

The task of thinking through the reasons for evaluation has implications for later tasks in the evaluation process. Questions concerning "why do it" will partly determine what is done and how it is done. For example, if the aim is to legitimate changes in personnel within the system, then a participatory approach to evaluative study ought to be used. If, on the other hand, the aim is to influence outsiders, then an outside consultant may also be very useful. The outside consultant, at least in the eyes of other outsiders, may be seen as capable of more objectivity than someone inside the organization, even though an insider might have more actual knowledge and understanding of the program.

## II. OVERVIEW OF THE EVALUATION PROCESS

A more detailed description of the evaluation process can be found in the I.R.S. report and in our working paper, "Guidelines and Criteria for Evaluating Vocational Rehabilitation Programs."

The comments below were written in response to an outline by another Task Force member which proposed that the evaluation process be discussed in terms of a series of steps or phases in the process:

1. Setting up objectives for evaluation
2. Selecting objectives to be measured
3. Choosing instruments and procedures
4. Selecting samples
5. Establishing measurement and observation schedules
6. Choosing analysis techniques
7. Drawing conclusions and recommendations

### A. Phases in the Evaluation Process

1. It should be explained that in practice the phases are not clearly differentiated or ever fully carried out. These phases are not mutually exclusive. There is much overlapping and a certain awareness of all phases must be possessed by the evaluator at all times. Evaluating, not to be confused with evaluative research, is really a rather sloppy process of continual reformulation of objectives, and criteria. This is how it is and this is how it should be.

2. In this step-wise approach, pre-evaluation study phase should be the first phase and implementation of evaluation findings should be the last phase. The contingency analysis is more fully explained in "What Needs to be Evaluated." In this analysis all phases are thought through in advance and problems and results anticipated. This thinking through takes place during the "selection of objectives" phase so that relevant objectives will be selected and the focus will be on things that can be changed. The implementation phases may create problems for personnel unless questions of how findings are to be utilized are also anticipated and resolved.

3. It may also be useful to discuss the evaluation process in terms of the following 5 functions:

a. Description (of the program and problem)

b. Understanding (analysis, data, etc.)

c. Judgment (decide on the value)

d. Recommendation (terminate, revise, etc.)

e. Implementation (make revisions, etc, planning, new goal setting)

Each function must be performed during the evaluation process if the process is to result in programmatic improvement.

The first function is that of simply describing what exists in the program: how many clients are served, how many resources are spent, what is the behavior of clients before and after they are served, what kinds of services are provided and by whom, etc.

The second function of evaluation involves trying to understand the relationship between the various behaviors and situations which have been described. This understanding is usually in terms of cause-and-effect. The approach or method for inputing causation can be either very formal

(as with control group and experimental design) or very casual (as when administrators simply assume that the observed change in client behavior is due solely to the services provided by their program).

The first two functions have received the greatest attention from researchers and academicians. Unfortunately, most discussions of evaluative research cease after the completion of the function we are calling "understanding." The rationale which is usually given is that the "scientific method" can not deal with questions of values and value judgments. While this statement is true, it does not necessarily follow that the evaluator's professional obligation should cease upon completion of the second function. He has further obligations to the organization which has sponsored the evaluation. It is just that he no longer possesses any unique expertise as he moves into the last three functions to be performed in the evaluation process.

The third function is that of making judgments concerning whether what has been described is good or bad. Ideally a knowledge of causal relationships has emerged from the performance of the previous functions in the evaluation process and influences the judgment of whether or not a program is performing adequately. Even though clients receiving services may be faring quite well in the employment market, it does not necessarily follow that the program is doing a good job. It may be that the program is having no impact at all, and that clients would do just as well without the program's services. In any case, the basis for deciding whether something which has been observed is good or bad is clearly dependent upon the values, goals, and expectations underlying the program.

The evaluation's criteria for judging worth should be made explicit, but in practice seldom are.*

The fourth and fifth functions to be performed in the evaluation process -- making action recommendations to the decisionmaker for improving the program, and implementing those actions which the policymaker or executive decides upon -- are usually ignored by most evaluation. It is important that the evaluator consider the requirements which these functions will impose from the very outset of the evaluation process. The evaluation study should be careful to consider not only whether current program performance is satisfactory, but also how the program could be improved. What kinds of actions or changes in personnel behavior, program operations, administration, or resource availability would improve program performance or correct weaknesses? Furthermore, evaluation study should give explicit attention to the probable effectiveness of the various policy instruments and forms of influence which the agency director or central office have available for trying to shift behavior at the field level in the directions which the evaluation study would recommend as desirable.

Are guidelines or decrees from the central office sufficient to modify personnel behavior at the field level? Will providing more case service funds necessarily improve the effectiveness of an office's handling of its caseloads, or result in higher quality vocational training and placement? Advice and information is thus needed not only on how well the program is doing, but also on what needs to be done to improve the program and how to accomplish what needs to be done.

---

*The nature of criteria for judging programs is described at greater length in Working Paper No. 173/RS003, "Guidelines and Criteria for Evaluating Vocational Rehabilitation Programs."

## B. Some Special Problems Concerning the Evaluation Process

### 1. The Limitations of Analysis

Analytic approaches are preferable to mere intuition. Even though the ideal models of experimental design and benefit-cost analysis can only be approximated, it is important as a referent in understanding the weaknesses of your particular methodology. Yet, analytic approaches should not be oversold or, at least, the barriers to successful utilization of analytic approaches should be made explicit. The glamour or appeal of some analytic approaches cause the weaknesses of such approaches to be overlooked. Even in approaches which use sophisticated hardware, error can result from bias or from improper design and utilization. Indeed, computer output and statistics can end up being used to conceal a poor understanding of the problem being analyzed.

Another barrier is that it's difficult to apply objective measures without certain resources, such as skilled personnel, time and money. These resources are frequently not available at many levels of the system. Beyond this, the results of past evaluative efforts in social programs and policies leave much to be desired. Many studies have been criticized for their methodological weaknesses, or the lack of significance of their findings or because unexplained biases are reflected in their selection of measures, analytical techniques, and experimental and control groups. Often program managers don't know how to interpret or to utilize findings. Thus, a meaningful evaluation relies not only on the choice of a proper approach, but also on the selection of relevant performance measures, the proper sampling and proper presentation of findings, etc.

2. Outcome Versus Process Evaluation

There are different approaches to conceptualizing evaluation and, although one approach may be more often utilized by rehabilitation than another, both should be mentioned. One concept may be termed the outcome or goal model and the other the process or systems model. (This is discussed by A. Etzioni, "Two Approaches to Organizational Analysis: A Critique and A Suggestion," Admin. Sci. Quarterly, May, 1960.) The goal model is reflected in the Task Force outline of seven phases in evaluations from setting objectives to drawing conclusions. This model determines a program's worth by measuring program outcomes in terms of certain specified objectives. The systems model includes the determination, description, understanding and evaluation of processes not directly related to the attainment of a specific objective. For example, in a systems model, the analyst might probe and use the distribution of resources throughout the organization, rather than simply assessing the final output of a program against the value of all the resources used within the program.

3. Conflicts in Values and Perspective

Another major problem is that evaluation activity is partly shaped by the values and intellectual perspective of the evaluator or sponsor. Different values can lead to difficult interpretations of problems and consequently, to different approaches to evaluation. Values among different actors in the rehabilitation system may conflict and it is important to understand such conflicts. For example, differences in values among counselors and clients may result in differences in understanding the cause of the problem which the client presents to the agency. The counselor may see the major problem as motivational, while the client may see the major problem as situational.

Similarly, when administrators find that programs are not working up to expectations, they tend to ascribe the reason to incompetence or malfeasance on the part of individuals, whether counselors, supervisors, or even legislators and superiors in executive super agencies. The evaluator, especially if he comes from a social science background, is more likely to ascribe failure to structural or situational problems -- poor training, lack of coordination, competitive social objectives, poor communication, limited resources, community structure, the state of the economy, etc. Even among evaluators, the kinds of explanation for program failure which will be stressed or even explored may vary depending upon the discipline of the researcher. Economists and sociologists ask very different questions about programs and, indeed, understand very differently how the world behaves.

Another example concerns views about the importance of work. Is work seen as a value in itself or instrumental in attaining something else, such as improved self-image or a "place" in society? How this question is resolved has implications for interpreting the objectives of vocational rehabilitation. Is a client "rehabilitated" when he is (a) fully employed; (b) fully employed in a job he is satisfied in; (c) able to care for himself, etc. According to the instrumental view of work, sufficiency in certain non-productive activities such as homemaking would be a valuable objective.

4. The Criterion Problem in Rehabilitation

A major challenge in rehabilitation is developing criteria for program success with clients which are broader than the simple 26 closure. Some states have considered (1) defining additional categories such as a quality rehabilitation ("Status 25") or a partial rehabilitation

("Status 27"), and progress along a number of vocational dimensions such as wage level, job stability, etc. Also, it would be useful to employ some nonvocational performance measures such as client capability for self-sufficiency (e.g. improvement in mobility, reduced need for attendant care) or even some additional measures which specify more fully improvement in vocational but unpaid activities (e.g. housekeeping, child care, volunteer work).

It is important to remember that when using aggregate indicators such as 26 closures, systematic (rather than eyeball) adjustments should be made for variations among states in casemix, local conditions, etc. Only in this way can comparisons across counselors, districts with states, or whole state programs be made.

## III. ISSUES AND CRITERIA IN EVALUATING REHABILITATION PROGRAMS

A third chapter in the original outline of the Prime Study was to discuss at length the framework of issues to be considered in program evaluation. The framework was to distinguish between evaluating case performance, case flow and overall service delivery, administrative and other support for field efforts at service delivery, and client and community impact. The chapter was to present examples of possible criteria in evaluating these various aspects of overall program performance, but was explicitly to forbear from making recommendations concerning desirable specific criteria or from even trying to present a complete list of alternative possible criteria. The Prime Study Group felt strongly that the selection of criteria was an important task which only a state agency, understanding its particular goals, constraints, and situation, could undertake. The Prime Study Group decided that it would not be possible or proper to try to advise state agencies on evaluation criteria. Our submission to the Prime Study as a suggested third chapter is represented by another working paper, "Guidelines and Criteria for Evaluating Vocational Rehabilitation Programs."

## IV. WHAT IS DONE IN PROGRAM EVALUATION?

### A. Alternative Approaches to Agency Conduct of Evaluation

What really takes place when someone evaluates a program? Sure, they look at whether the program is achieving its objectives, whether it could be operated more efficiently, how things might be improved, whether what is being achieved is worthwhile. But how do they actually go about making such determinations?

On the surface, there appear to be four main ways of going about evaluation. When each is looked at more closely, they often amount to the same thing. The four general lines of attack are as follows:

#### 1. Routine Monitoring of Program Data

Someone looks at the data which is routinely generated during the program, assembles the data in some way that appears to be appropriate for judging how well the program is doing, compares the reassembled data with rules of thumb or norms taken from the past history of the program, and says that things are going well or badly. Such data need not necessarily be program statistics. Rather, data could be treated more generally as information. Thus, a supervisor routinely reviewing a counselor's records is performing this kind of evaluation, even though he is not looking at statistics per se. Rather he looks to see if the kinds of services being given look appropriate, if the counselor has recorded his diagnosis of client's needs and if that diagnosis looks reasonable, if the client seems to be progressing in a reasonable way.

#### 2. Periodic Speical Reviews by Inside Staff

A special review is undertaken by central office staff or by a district administrator, supervisor, or even the counselor of particular

problems. The person doing the evaluation goes into the field or calls up individuals who should have the information necessary to make the evaluation, asks the questions or observes behavior, and then makes an evaluation judgment. What distinguishes this line of attack from the earlier line discussed is that the questions which were asked were not routine. Special questions were asked; one looked for something which one would not normally look for. The evaluation was generally a one-shot affair. It might well be repeated on an annual basis, but it seldom is conducted on a more regular basis. Also, the evaluator was fundamentally eyeballing the situation and then making his judgment.

3. Obtaining Assessments from Informed and/or Objective Sources

The third line of attack which agencies frequently undertake is to simply ask the opinions of others as to how well the program is operating. The opinions are taken at face value. The agency may want to know why the person has the perception or opinion which he does, but generally what is most important is that someone "who ought to know' thinks the program is going well or badly. Who is asked for such opinions? Who is ascribed as someone "who ought to know?" Sometimes a central office will ask a supervisor, or a supervisor will ask a counselor. Such people are asked when they are thought to be sufficiently informed about the issue of concern, when it is believed that they have values and contextual knowledge sufficient for making an acceptable judgment, and thus when they are not perceived as having a vested stake or self-interest in the evaluation being made. Often on fundamental issues of how well the program is doing, the agency is aware that no one in the agency may be sufficiently objective. At this point, the agency turns to "outsiders" -- "experts" in the field usually. Such people are assumed to have as much

if not more knowledge than the staff employed in the program. (Often the staff really believes that the experts actually have less knowledge than they do themselves.) The experts do possess objectivity, at least they appear not to have a self-interest in the results of their evaluation and their opinion is seen as legitimate. Sometimes, when executive departments or legislatures rather than the agency have required the evaluation, even "experts" in the field are viewed with suspicion and the opinions are asked of auditors or generalists who lack previous experience with the program. In such cases, agencies are always sure to insist that the evaluator be very explicit about the basis for his judgment, indeed that he prove his conclusions.

What is most interesting, however, is that in assessing how well programs are doing, one seldom asks the client -- even though he presumably is the person who is supposed to be the subject of the program's concern and changes in his behavior and capabilities are the goal of the program.

4. Special Program Analyses or Evaluation Studies and Research

The fourth approach is more formal. Someone undertakes a formal evaluation study. The study seeks to be "objective," that is, to present conclusions which are clearly derived from facts and which are not subject to manipulation or whim of any set of individuals. To achieve this goal, the evaluator turns to "scientific method" and the model of "experimental design" in research. In practice such evaluations never conform as closely as even the evaluator might initially have hoped to true experimental design. Also, the studies do not ultimately escape the intrusion of the evaluator's values. Such studies always ultimately require the use of judgment in the interpretation of what is fact and in determining what

the implications of agreed-upon facts are. The study does at least attempt to be "scientific," however. That is, the study attempts to develop findings (or test hypotheses) through a clearly articulated methodology and presents the findings and methodology in such a way that the study process could be precisely repeated by another analyst or evaluator and the findings thus reconfirmed or validated.

5. Similarities and Differences Between Approaches

All four approaches or lines of attack have several things in common. All approaches require that someone describe a situation, that is to describe what exists or what has happened. All approaches at some point require that the evaluator have an understanding of how things normally go on, so that he has a context in which to understand and judge the meaning of the particular situation which he has observed or measured and then described. (NOTE: Sometimes an understanding of how things normally go on is part of the goals of evaluation, as in the approach of systems analysis to evaluation.) All four approaches require at some point that the evaluator judge whether what he describes is good or bad. The ability to make such a judgment requires that the evaluator have some image of how the situation ought to be. Social scientists often like to call such images and understandings "models." The image used in judging is a normative model. The understanding of how things normally go on could be called a descriptive model.

Most of the time when evaluation is done internally within an agency, the criterion of how things ought to be is that the situation being evaluated should be consistent with how things normally go on. One uses average program performance at any point in time or historical program performance or rules of thumb or the average performance

nationally as the basis for judging whether a situation is good or bad.

Once a judgment is made, two additional steps are involved in an evaluation effort in all of the above lines of attack which are often overlooked by academics and the evaluation literature. First, one must decide what program actions or decisions should be undertaken on the basis of the judgment which has been made. Often evaluators leave this to the "policymaker," making no specific recommendations themselves. This tactic does not eliminate the need for someone (i.e. the Director or his staff) to formulate conceptually what needs to be done on the basis of the description, understanding, and judgments which have been produced during the study phase of the evaluation. Second, once it has been decided what needs to be done, it remains for someone to implement the specified program actions and decisions.

What differs among the four approaches at first glance appears significant, but on a second look seems inconsequential. The first key difference is the reliance on formal or informal evaluation approaches, or perhaps more appropriately the extent to which the approaches strive toward objectivity or scientific "validity." Only the last approach is truly formal. Yet the first approach to the extent that it manipulates routinely published statistics tends toward formality. Similarly, the third approach, when outside non-experts (e.g. auditors, evaluation consultants) are drawn on for the evaluation approach, is often forced into formal evaluation methods before the opinions of the evaluator will be accepted. Second, the approaches differ fundamentally with regard to who is the evaluator. The first and second approaches draw on insiders to the agency system, the third approach can draw on insiders or outsiders,

the fourth approach seeks an evaluation process which rises above the issue of who is the evaluator. In each approach, however, the evaluator must draw on some normative criterion and descriptive model of how the system functions before he can make his final judgment. This judgment criterion and descriptive model often is similar across approaches, regardless of who the evaluator may be.

Which approach is proper in any given situation depends on who wants the results of the evaluation. If the evaluation is needed to persuade outsiders of the legitimacy of program demands for more resources or for maintenance of current funding, then the third approach where the assessment of outside experts, clients, or other evaluators is sought, or the fourth approach where the evaluation appears to be based on a scientific research design is most appropriate. If the evaluation is really being sponsored for internal purposes, in order to identify ways in which the program can be improved, then the first and second approach may often be sufficient. Available resources of money, time, and staff skills for conducting evaluation efforts also must influence the choice of an evaluation approach. What must be stressed is that there is nothing intrinsically less proper about the first two approaches than the latter two approaches. Depending on who wants the results of the evaluation, the first two approaches may often be less costly and as accurate and thus, more efficient and useful ways of attack than the more formal 'scientific' approach.

The problem with the first two approaches and also the third approach is that the evaluator and those who receive the evaluator's conclusions must be forced to remain sensitive to where the evaluation approach can result in false inferences about what the impact of the program

really is. There are two practices which could be required of evaluators by which this sensitivity or awareness can be maintained. First, the evaluator should constantly compare his approach to the classic model of experimental design and consider what biases and inferences may result from the way in which he has tried to describe and impute causality to the program he is evaluating. Second, the evaluator must be obligated to be explicit about presenting the model which he is using to understand the program which he is describing and the value criteria which he is using to judge what he has described. Similarly, he must be explicit about what he has looked at as he has tried to describe the impact of the program. If this explicitness is achieved, it will be possible for those within and outside the program -- regardless of level in the agency or perspective -- to judge the reasonableness and validity of the methods and findings of the evaluation study.

## B. Some Suggestions for Avoiding Pitfalls in Evaluating Vocational Rehabilitation Programs

In the sections which follow, we discuss first some general suggestions on problems to be encountered in conducting evaluation. Then we shall discuss at length problems which result as evaluation studies or efforts deviate from the classic model of experimental design. Finally, we shall give a few comments on sources of data. The outline of the "What is Done" section which was developed by the Prime Study Group in its initial meetings focused exclusively on sources of data rather than actually on what is done with the data.

### 1. General Comments on Biases and False Inferences

Be conscious of biases produced in sampling the client population in order to assess how well the program is doing generally with

specific kinds of clients or in order to project the program's likely performance in the future as services are expanded to more and perhaps different populations. Often the clients you are best able to reach via mail and phone contacts are the clients you have had the most impact upon. Judging the program on the basis of the experience of such clients overstates the program's achievements. Similarly, as the program expands in the future, new clients may have more intractible problems and difficulties, even though they are formally classified as similar in disability to clients served in the past. Economists and businessmen distinguish between average costs and marginal costs. Usually marginal costs, that is, the costs of producing one more unit of output or of treating one more client, are much higher than the average unit costs for all production or the average client costs for a given disability. Thus, the costs of serving additional clients in the future may be higher than current costs. (NOTE: This has nothing to do with inflation whatsoever.) Again, if new clients to be served in the future are fundamentally different in kind from the average client currently served, one should be careful to avoid projecting future costs and success rates on the basis of current experience in the overall program. As rehabilitation programs, for example, begin to emphasize services to public assistance recipients, costs and success rates may dramatically diverge from past experience.

It is also difficult to make comparisons of the performance among counselor, districts, and states unless the performance measures are first adjusted (or "normalized") for case mix. Even then comparison must consider varying local conditions: the state of the local economy, prevalence rates and the structure of need in the community, community

infrastructure and the availability of services through other agencies or programs (e.g., Medicare, welfare, the voluntary sector), etc.

### 2. Comparing Evaluation Study Designs with the Classic Model of Experimental Design

Although the classic experimental design is the "ideal model" for conducting an evaluation, in the sense of measuring the impact on clients causally produced by the program, the resources and instruments for conducting such research rarely exist. Moreover, simpler analytical and investigation techniques can produce insights just as useful, even though they are less "scientifically" rigorous. Yet, it is important to be aware of how your own evaluation design has deviated from the classic experimental design in order to understand where your design may result in false inferences about program performance and needs.

According to Edward Suchman (Evaluative Research, N.Y., 1967) the ideal or "true" experimental design involves a set of procedures such as are illustrated in the accompanying flow chart. We shall discuss these procedures in turn, pointing out the pitfalls which can await the evaluator of rehabilitation programs as he is forced to deviate from the experimental design model.

a. Definitions of the Target Population. This is no easy task. For example, should the target population for rehabilitation programs be all the disabled and in need of services, when the programs lack resources to serve more than a fraction of that population? One might question why false expectations are raised on the part of clients and the public is allowed to sit self-satisfied at having met the need of "the unfortunate handicapped." Would it not be better for the program to state frankly the size of the population it is able to serve with its

resources and to let the public become aware of the size of the unmet need in the handicapped population? Similarly, should the agency evaluate its performance on even the basis of how well it serves the total referred population? Many referrals exist on paper only: the handicapped person has never walked through the doors of VR offices. Given too few resources to serve all those who do walk through the door and given that a potential client does not show enough initiative to apply or appeal for services, why count him or her among the agency's target population? Also, many of those referred when questioned and reviewed show no need or disability qualifying them for services or alternatively show such severe disability that rehabilitation is not feasible. Why should such clients be considered part of the target population of rehabilitation programs? In short, the choice of the definition of the target population of the program involves fundamental questions of values, political and managerial judgment, and program objectives.

If one does accept that entire handicapped population in need of services is the proper target of the agency, one still has difficulty even measuring the size of, much less locating that target population. Generally, national data on prevalence is available, but little is known about how to make adjustments for particular states or areas within states in estimating prevalence given varying rates of urbanization (urban/rural/suburban), sizes of minority populations, industrial and occupational structures, etc. Although these varying conditions appear to be associated with much higher or lower prevalence rates of disability, few states adjust estimates of a real need for services on the basis of such conditions.

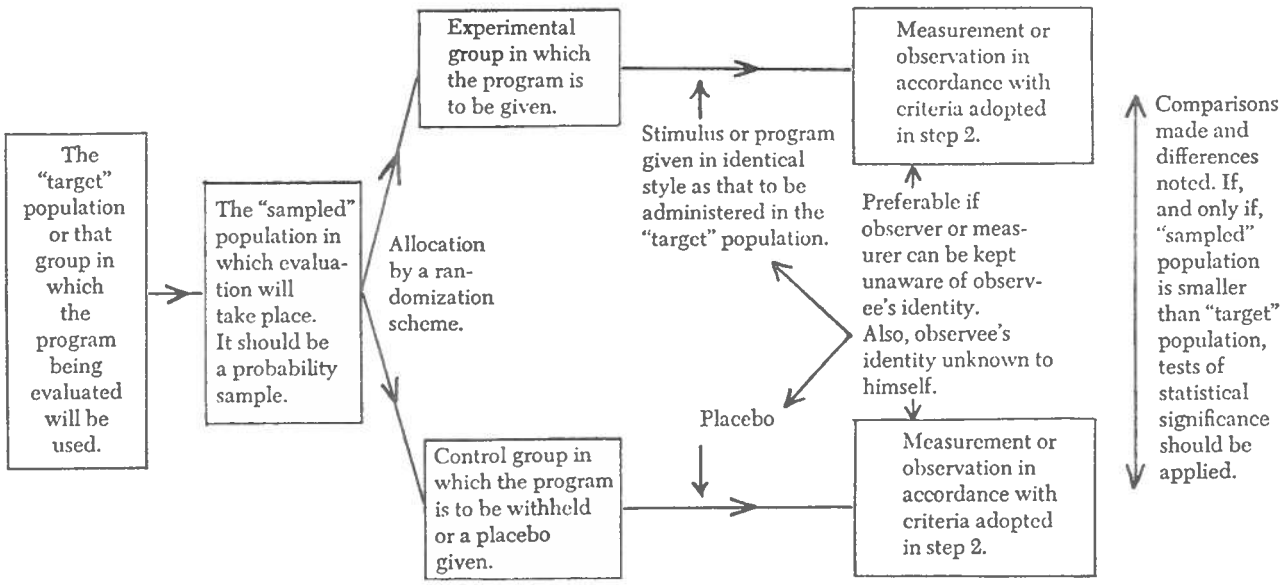
b. Drawing a Representative Sample. Be aware of potential biases which could arise from sampling strategies. Examples of when misinferences can arise include:

i. Judging the effectiveness of a program from demonstrations involving volunteer clients.

ii. Estimating past or probable future program success with a particular subpopulation from past experience with either (a) the entire client population served by the program, or (b) a subpopulation served in the past which differs in important ways (e.g., demographic and disability characteristics, income level, location) from the subpopulation of particular interest.

iii. Sampling clients during a particular time period for measuring program effectiveness over the entire fiscal year. For example, if in some states, there is substantial examination and closure of cases at the end of the fiscal year, reviewing program performance via a sample drawn from January closures can be very misleading. Agencies need to consider whether there are variations in the volume and type of case closures over the course of the fiscal year which require adjustments in any evaluation design similar to the seasonal adjustments made in assessing agricultural and industrial production.

iv. Taking samples on a convenience basis, e.g., those clients still living at addresses in the files, who have phones, or who are willing to respond to mailed questionnaires. Careful consideration should be given to whether the nonrespondents or the clients who cannot be located differ in fundamental ways from those in the sample. If they do, then the inferences drawn from the responding sample populations may be misleading. In 1936, one of the biggest landslide elections in




FLOW CHART

This flow chart illustrates optimum principles and sequence to be followed in conducting a valid experimental design to evaluate a health program. (Reproduced from Greenberg, Bernard G., and Berwyn F. Mattison, "The Whys and Wherefores of Program Evaluation," *Canadian Journal of Public Health*, vol. 46, July, 1955, p. 298.)

Presidential history, one pollster for a national publication predicted that Alf Landon would easily beat F.D.R. on the basis of a sample of voters drawn from the phone book. The pollster simply forgot that a large share of the population -- especially the poor and working man -- did not have phones at that time.

Quite apart from noting biases that can emerge in drawing a sample, one should also remember that 100% samples are no solution in avoiding biases and indeed, are very wasteful. A 10% sample can usually yield information as valuable and acceptably accurate as a 100% sample, and yet much less costs are absorbed in terms of counselor, administrative staff, and client time, data processing, etc. Indeed, a 1% or 5% sample can often be sufficient to answer the questions of administrators. Historically, our problem in rehabilitation programs has been that the men in Washington who arrange for our collection of "statistics" are not statisticians, but bookkeepers. Yet, the information that is sufficient for accounting our program to legislators and auditors is different in kind and magnitude from the information needed for management and planning. There is no need to collect both kinds of information on a 100% sample basis. In future reviews of the R-300 data form at the Federal level, this distinction between accounting and management should be kept in mind, and serious consideration should be given to reducing the length of the compulsory R-300 form for the 100% client population and to changing and even (if necessary) lengthening the form to meet the needs of management but collecting such management-oriented information on a sampling and temporary basis.

c. Allocation of Clients Into Experimental and Control Groups. When Budget Bureau types and economists review programs, they keep asking the nasty questions. They want to know not only what happened to the

clients you served, how many clients did you serve, but also they sometimes have the nerve to ask "how do you know that your program was really responsible for the client getting the job?" The challenge may be made: "What would happen to that client had he never gone to your program in the first place? How do you know he wouldn't have gotten the same job in time anyway?"

In determining the answer to such questions, the social scientist using the experimental design for evaluation points to the experience of a "control group." That is, he compares what happens to clients who received rehabilitation services and whose cases were closed as successfully rehabilitated with what happens to some other population group whose characteristics are similar to the group served. What differences exist between the experience of the two groups are imputed to have been caused by the services which one group received and the other group did not receive.

Putting together control and experimental groups on paper looks straight-forward, at least to the academic. Experimental design calls for taking the sample of the target population, matching them in terms of characteristics, and then randomly allocating them to an experimental group and a control group. The experimental group receives services, while the control group does not receive services or receives instead a "placebo," that is, something which the client thinks are services but which are not the services which the program wants to evaluate. (The reason for giving "placebos" is to guard against the morale effect that can occur when participants in an experimental group or the control group, temporarily respond in atypical ways because of the excitement or frustration that they feel as a result of knowing that they are in an

experiment.) Where matching clients in terms of their characteristics is not feasible prior to their allocation to experimental and control groups, the evaluator must normalize the population after the experiment before comparing the experience of the two groups. That is, he must adjust the populations statistically so that he controls for differences between the populations which existed at the outset of the experiment. Such adjustment is readily possible through analysis of covariance and other statistical techniques.

Unfortunately, carrying off such an experimental design is much more difficult in practice. Society will not usually look kindly on openly refusing services to clients for the sake of experimentation, even though society claims to want programs evaluated. (Some have noted the irony of such social disdain for refusal of services for the sake of experimentation. Since we lack enough resources to serve all clients and must thus reject some clients anyway, why does society have qualms about handling this necessary rejection activity in a way which increases our scientific knowledge about the effectiveness of services?) Many evaluation specialists, recognizing the difficulties of constructing pure control groups, are now suggesting that agencies instead concentrate on comparing program strategies (where two similar groups of clients receive different kinds of services or receive similar services through different kinds of programs), rather than trying to compare clients receiving services with individuals who have not received services. This approach might well be recommended to state rehabilitation agencies. Quite apart from the difficulty of refusing services to a control group, it is also administratively too complex to randomly allocate clients to experimental and control groups and indeed it is usually impossible to find in a given locale enough clients with similar characteristics to form perfectly matched pairs.

Given these difficulties, some research groups have tried to use other kinds of client populations as control groups. Sometimes, when the biases which were inherent in the use of such control groups were not recognized by the researcher, disastrous inferences were made about the performance of rehabilitation programs. For example, in one sizable study recently commissioned by D/HEW,* the research group chose to use a sample of 28 and 30 closures as a control group for comparison with a sample of 26 closures. The researchers discovered that 60% of the control group had achieved employment success comparable with the clients whom Vocational Rehabilitation agencies had declared rehabilitated. The researchers thus wrote off 60% of the originally measured impact of the rehabilitation program as not really due causally to the rehabilitation program. As a result, the benefits of the rehabilitation program no longer significantly exceeded the costs of the program, and the researchers concluded that the program was not a notable success and should only be expanded if considerably modified. The mistakes of the researchers in using 28 and 30 closures as a control group are conceptually straight-forward:

i. The control group differed substantially from the experimental group (26 closures) in predictable demographic and disability characteristics, but the researcher made no attempt to normalize these differences before comparing the employment experience of the two groups.

ii. The control group internally was actually comprised of two very different populations both of which significantly differed from the experimental group. Individuals closed before plan (30 closures) were less severely disabled than the group of successful rehabilitants. They often left the program precisely because they were able to obtain

---

* Institute for Interdisciplinary Studies, Toward an Analysis of Federal Rehabilitation Policy, Minneapolis, 1971.

employment through their own initiative and skills. Comparing 26 closures with 30 closures thus significantly understates the change in client employability due to the rehabilitation program. (This problem with early dropouts is commonly experienced with all manpower programs. See Thomas Glennan's RAND monograph, "Evaluating Federal Manpower Programs.") In contrast, clients whose cases are closed after they have received considerable services but who have nonetheless not obtained employment often tend to be more severely disabled or to belong to subpopulations (women, minorities) which generally experience difficulty in obtaining employment. Even after normalization, comparing 26 closures with 28 closures thus overstates the change in client employability attributable to the rehabilitation program.

iii. At least some part of the gains and improvements experienced by individuals recorded as 28 and 30 closures might well be attributable to the services received from the rehabilitation program. In short, the control group in this case partly shares in the benefits of rehabilitation programs, rather than represents a zero baseline against which the benefits of the programs can be calculated.

The research study in question failed to note any of these distinctions and as a result arrived at evaluation findings which were very misleading. (Unfortunately, wrong inferences from research may nevertheless have as much political and managerial impact as correct inferences.) It is thus clear that being constantly aware of the biases which are introduced via deviations from true experimental design is critical to the conduct of a good program evaluation.

Our point here is not that only a pure "scientific" experimental design is acceptable, but rather that an evaluator must constantly keep questioning where he may be making false inferences about the program.

One excellent way by which to discover where inferences may be incorrect is to carefully think through how the evaluation deviated from the experimental design and thus what kind of biases may have crept in. Evaluators will always make compromises, indeed severe compromises, with the scientist's experimental design. Agencies cannot afford annually to think about control groups or even collect extensive follow-up data on clients. Such in-depth program evaluations are expensive and should only be undertaken every few years. It is possible, on the other hand, to routinize the collection of follow-up data on a sample of client closures on an annual basis. This would be especially inexpensive if it were possible to assess the Social Security earnings records of clients. In comparing clients, however, the evaluator -- as described above -- must keep in mind demographic, disability, and locational differences between clients and between closure statuses.

d. Administration of the Program or Treatment. The actual program or "rehabilitation service" given a client may differ from what is written on paper. Thus, it is important to observe and record what actually occurs in the program. In assessing program effectiveness, one should analyze the impact on clients of variations in the kinds of services given (service mix), in the training of the counselors working with clients, etc.

Many things that happen during the program are not readily explored with numbers or measures, e.g., the type and amount of counselor supervision, counselor morale, coordination with other programs in the community, the kinds of supplemental services available to clients in the community. Observations, however judgmental, should be recorded on such factors and used in analyzing and interpreting data on clients, counselors, and services.

(Such observations can, in fact, be handled analytically via nominal and ordinal variables with contingency tables and dummy variables in multivariable statistical analyses but we'll leave such concerns to the analyst.) These judgmental observations can be especially useful in figuring out what kinds of management actions may be able to improve program performance after the evaluation results are in.

As discussed before, it is desirable but difficult to find a "placebo" which could be administered the client population being used as a control group. It is also often not feasible to let some needy potential clients go without service. Due to these difficulties it may be necessary to forego the use of traditional control groups. Instead, rely on design in which programs or strategies within rehabilitation are compared, such as manpower training and the traditional counselor-client approach. Alternatively, compare the client impact of the rehabilitation program with other programs sponsored by Welfare of D/Labor (e.g., WIN). In such comparisons, it will still be critical to normalize the client populations as much as possible.

e. Measuring Program Impact, Analysis and Comparison. At the outset of any evaluation, the evaluator with the program administrator must determine what they want to look at to determine if the program is meeting its objectives and what criteria they will use in determining whether a given level of performance is acceptable or not. In measuring program impact, it is important that the instrument which is used give consistent results regardless of who administers the instrument. Thus, using 26 closures as a measure of success can be misleading if counselors interpret in different ways what kinds of client achievement justify closure as a 26. Similarly, it is important that the measure of program impact which is used validly measure the objectives of the program.

Thus, it is more appropriate to use actual data on client employment, income, and capability to assess client impact than the proxy of whether or not the client has been closed in the 26 category. Proxies are often necessary because of cost constraints and time limitations. Periodically, however, the agency should re-examine the proxy to see if it is applied consistently across districts and counselors and if it validly represents a real improvement in client functioning and employability. Routine follow-up studies for samples of client population are feasible. Such studies should embrace not just 26 closures but other kinds of closures as well. The agency should seek to identify delayed employment gains and non-employment benefits which result from its services.

Quite apart from re-examining the legitimacy of the 26 closure as a proxy for rehabilitation, it would be useful for agencies to conduct studies periodically on the validity of R-300 data. The R-300 data, after all, is the key source of management information on what is happening in the program. This validity check could be done through follow-up studies with the client and with other agencies (checking referrals, welfare payments, etc.) or through review of the original case folders. In this way, it can be determined not only whether the proper closure category was used, but whether the impact of the program is being under- or over-estimated. While some clients classified as 26 closures (perhaps as many as 15-25%) will be found unemployed or as having received little tangible service in such a review, the agency should not feel shocked. Even after adjusting for such follow-up results, the overall program performance is still far better than most other social service and manpower programs. The overall cost-benefit return still remains extraordinarily high. Moreover, such follow-up studies will also often discover that clients in other closure statuses have made delayed

improvements in their employment situation or in their capability for non-employment activities. In many cases, these gains can be attributed directly to services received from the rehabilitation program.

This discussion does not mean to imply that the evaluator should look only at statistically significant findings with perfectly valid data, nor that those aspects which cannot be quantified should be ignored, nor that if the evaluator sees something new when he or she is in the field that wasn't anticipated in the design of the study or the measurement instruments, the observation can't be included in the evaluation effort.

The use of good field observations and common sense judgment in data analysis is important. The counselor in a district office who is sloppiest at record keeping may in fact be the most effective counselor. How can the evaluator determine if this is so? Research may provide some clues for such a determination but the best guide is good professional and peer observation and judgment. Similarly, the analyst in having put together his statistics must still ask himself what confidence level should I appropriately use to reject the hypothesis that the differences between the experimental and control groups which I have observed are not random, and arisen from pure chance. Judgment is inevitable in interpreting and understanding what one sees and measures. Indeed, even in the best evaluative research study, even using an experimental design, when the analyst puts together the last statistic and the field interviewer presents his final descriptive observation, it still may not be clear if what has been observed and measured is "good" or "bad," or whether and what kinds of management and program actions seem to be justified by what has been determined to be the impact of the program. Such judgments depend on the context and do not follow directly from the

research. There is no black box in evaluation studies, however "scientifically" conducted, which lets the evaluator and policymaker off the hook from making judgments and decisions amidst uncertainty.

3. Assorted Comments and Suggestions on Data Sources

The comments which follow were posed as suggestions to the task force members who were drafting the chapters on data sources for evaluation studies. The comments suggest additional or supplemental points which had not previously been discussed or emphasized in the meetings of the Prime Study Group.

a. Using an Advisory Board composed of former clients, disabled people and other citizens can be very helpful in suggesting ways to improve program design. Yet such advisory groups have seldom been exploited as a technique by state agencies. Much more needs to be done in this area. Similarly, the suggestions of counselors and field personnel also need to be routinely introduced to the evaluation unit via some formal mechanism or process.

b. Both of these kinds of Boards or groups -- clients and counselors -- can be very useful as well in suggesting hypotheses, reviewing questionnaires, interpreting the programmatic meaning of data findings, and reviewing the conclusions of an evaluation study and formulating recommendations. Central office staff, even when they have come up through the counseling ranks, often become removed from the realities of day-to-day field operations. They have limited ability to identify the reasons behind what appear to be failures in performance at the local level, and thus experience difficulty in formulating useful program recommendations. They may not understand the costs which are involved at

the field level in completing extended questionnaires. They often find it difficult to formulate questions in words which both can be understood by field personnel and yet precisely convey what the evaluator wishes to know.

c. When an evaluation unit uses studies which have been conducted in other programs or in other state agencies, they should be careful to ask several questions: (1) Is the study's conclusions generalizable to the state program of concern to the evaluation unit? (2) Are the definitions used for collecting data (e.g., the R-300) in the other state comparable to those used in the unit's own state? (3) What were the special assumptions which were made by the "other study" in formulating its research design and deriving its conclusions, and are these assumptions valid for the unit's own state program? Often, as with the cost-benefit studies which have been conducted by various state rehabilitation agencies, the assumptions used by various studies are quite different and non-comparable. Since many state agencies would like to be able to compare the relative success of their programs, they should be encouraged to use similar assumptions and study designs whenever feasible.

d. When collecting data on clients for program evaluation, it would be more economical and just as accurate if state agencies were to collect data for samples of the client population, rather than for the whole population. This approach, if pursued intelligently, could allow the R-300 forms which are collected for every client to be reduced in length to the minimum amount of information needed for administrative functions. The purposes of auditors and evaluators can be served by appending supplemental questionnaires to the basic R-300 documents which were designed for investigating particular problems, evaluating program

impact or effort, or conducting basic and applied research. Such supplemental questionnaires would only need to be administered to a sample of clients, however. Even now some state agencies require supplemental questionnaires to be completed in addition to the R-300 instruments; these state agencies usually require that the supplemental questionnaires be completed for all clients, however, thus greatly expanding the paperwork load placed upon field personnel. A more flexible strategy of sampling and periodic questionnaires would reduce the paperwork within agencies while at the same time greatly increasing the information available to the central office and facilitating the work of program evaluators.

e. In manipulating R-300 data, the federal government should cease publishing performance statistics for state agencies (such as the number of 26 closures per 100,000 population or average expenditure per 26 closure) unless they also publish comparable normalized statistics which adjust for disability or case mix and which reflect the wide variance across states in service mix and program emphasis. R.S.A. admits that it is not proper to compare the performance of state agencies with such unadjusted statistics, but continues to publish the statistics. Once the unadjusted statistics are available, as the only performance data available, they are picked up and used by legislators, D/HEW administrators, and even state agency directors to compare state programs, in spite of the invalidity of such use.

f. Finally, the so-called "clinical" or "planned change" approach to evaluation should be mentioned. This approach, which has long been employed in the rehabilitation profession, can be used not only as counselors work with clients but also as program evaluators work with organizational systems. The model for this approach to organizational

problem-solving is similar to the familiar model used in most helping professions. That is, a worker enters into a collaborative relationship with a client or client group and the relationship, itself, which is used for diagnosis and treatment or training, facilitates desired change. The path toward change is deliberate and is mutually agreed upon. In this approach the helping person is referred to as a "change agent." The change agent is typically a behavioral or social scientist who evaluates organizational effectiveness in terms of interorganizational relationships.

Although the focus of planned change, and its resulting solutions are on a micro rather than a macro level, this approach is particularly useful in confronting age old problems which are hard to quantify, such as human collaboration and conflict, control and leadership, and utilization of human resources. Further, it is one of the few organizational change models which has a framework for going beyond the stage of mere identification of the problem and on to the stage of implementing the results of research.

The term "clinical" is also used to refer to another approach to organizational evaluation. An in depth case study assesses, in clinical detail, certain organizational variables from the perspective of particular actors in the system, such as clients, counselors and/or administrators. Although case studies can be subjective and limited in representativeness, such studies can be useful in formulating hypotheses and in highlighting important aspects of a problem which warrant more extensive investigation.

## V. WHO SHOULD DO PROGRAM EVALUATION?

Generally, the comments which follow concern when or how insiders, outsiders, the Federal Government and consumers should be involved in evaluations.

All levels with the state program -- district administrators, counselors, supervisors, and even consumer representatives -- should be involved in suggesting program elements for evaluation, reviewing designs and questionnaires, and reviewing and evaluating the findings and interpretation of findings in evaluation studies. Such field staff involvement may provide the evaluation staff with insights about program operations, client needs, and problems of which it had previously been unaware. The involvement also facilitates later utilization of research. Field staff and consumers are especially resistent to long questionnaires, poorly worded or naive questions, and research intrusions on their time which can not be justified in terms of probable value of the findings for improving the program. In thinking about who should be included, it may be useful to think in terms of who is now excluded, and then to consider ways to involve this group.

### A. The Federal Role

Generally, the federal level should not impose evaluation, but should assist states financially when evaluation costs to states would be prohibitive but only at the request of states. Federal government assistance to state agencies should be considered in the following areas:

1. Subsidizing the initial start-up costs for hiring staff and creating a capability in state agencies for evaluating and analyzing rehabilitation programs. Technical assistance should also be made available on request.

2. Subsidizing the training of staff skilled in program evaluation via both a) in-service training and b) basic graduate school training.

3. Developing computer and analytical software models, for state use in evaluation, such as cost-benefit analysis, payback period estimation, estimating performance measures adjusting for can mix, etc.

4. Assisting states in developing and promulgating common assumptions and definitions of performance measures for judging program performance. Legislatures and outside critics often cite other states' performance in critiquing a state's rehab program, without recognizing the different definitions underlying the reported statistics, (e.g., cost-benefit) and the differences in local conditions.

5. Developing demonstration programs where a common evaluation design is used from the outset across projects. The current piecemeal demonstration efforts impedes learning from such efforts and wastes money by forcing each demonstration project to develop its own evaluation designs. If a basic evaluation study design were to be developed, made available to state agencies, and implemented, the data generated across projects would be comparable and a more complete, generalizable and powerful assessment could be made of the program strategies or concepts being demonstrated. This use of common study designs could be further

linked with planned experimentation with new techniques and program strategies. As rehabilitation agencies enter new areas such as drug abuse and alcoholism, such experimentation could be quite useful.

6. Performing data processing. Small states cannot afford expensive computer systems. Rehabilitation agencies in large states often have low priority in access to state computer networks. A feasibility study might be done of having Federal regional offices do processing of R-300 and other data for state agency evaluation efforts. Indeed, it might even be appropriate (though radical) to explore use by smaller state agencies of regional computer capabilities for routine processing of R-300 data for ongoing program management. States could set their own specification for data input and output formats. Given the large economics of scale in M.I.S. development and computer technology, such a Federal service upon state demand might save state agencies considerable money as well as making possible improved management information systems and evaluation efforts.

B. Insiders

Evaluation personnel inside the agency should be used for routine management and efficiency evaluations, for on-going program assessments, and particularly when the objective of the evaluation is agency change. Also, an insider should be assigned to work closely on evaluation whenever an outsider is used. In this way the outsider will have access to relevant information about the agency and there is increased likelihood that evaluative findings will be utilized. The ideal evaluation staff unit should contain people possessing among them graduate-level skills in the application of experimental design, statistics, and economic analysis.

At least one member of the unit should have extensive practical field experience in the rehabilitation agency at the counseling level. All members of the staff would ideally also be skilled in inter-personal relationships. Academic recluses make very poor and ineffective evaluators, no matter how good their credentials and substantive knowledge. The evaluators should be routinely mixing with line administrators and field staff if their work is to be effective. It would also be desirable if at least one member of the evaluation staff had a knowledge of organization theory, planned social change, and public administration.

One excellent suggestion that has been employed in private corporations and in OEO is to rotate on an annual basis one field person with the evaluation unit. Because of natural reluctance to pick up and leave a family, clients, etc., the individual selected for the rotating assignment is often either a young counselor or a new addition to agency personnel coming from another state's program or even just out of graduate school. Such a strategy is an excellent way to bring new insight into current evaluation efforts, to socialize personnel to broader management and policy issues, to establish a network of contacts for the evaluation unit around the states for exploitation in future years, and to place eventually into the field counselors and supervisors who will understand the needs of evaluation and who will act to inform and change the perspective of other field staff, who may tend to resist change and innovation.

C. Outsiders

Outsiders are typically brought in to conduct an evaluation because an outsider is thought to be more "objective." But, because outsiders may hold biases, it is necessary to define what is meant by "objectivity."

The outsider (or insider) must be required to make explicit:

1. the criteria he uses for judgment;

2. the basis for rendering his value judgment. In this way his observations become replicable and his conclusions can be more properly assessed by the administrator.

An outsider should be used:

1. when the agency staff cannot control its biases by making them explicit;

2. when agency staff lacks the needed technical skills for evaluation, or

3. when a skill is needed only for a brief time so that the hiring of full-time or permanent staff members is not economical. Outsiders contracted for evaluation studies should also have skills in experimental design, statistics and economic analysis.

D. Consumers

Consumers should be routinely involved in identifying unmet needs and in determining and valuing the various types of impact which their experience with rehabilitation has produced. For example, state agencies might occasionally sample client judgment as to whether the client has experienced an improvement in his capability and welfare along a number of dimensions. Employment impacts represent only one dimension. (The Oklahoma agency will be attempting such a study as part of a recently funded grant.) Also impacts can be imputed from changes in consumer-reported behavior. One need not elicit only subjective judgments from the consumer such as whether the client "liked" the counselor, or thought that the rehab services he received were instrumental in his obtaining or advancing on his job.

If a program is sponsored jointly with another agency then it would be useful to sponsor a joint evaluation. There is a greater likelihood that effective joint action will result if the need for and type of action are identified by both agencies in a mutual evaluation effort.

## VI. HOW TO ORGANIZE EVALUATION ACTIVITY WITHIN AGENCIES

### A. Organizing an Evaluation Cycle Within an Agency

It is important to develop mechanisms by which all levels of the rehabilitation system are brought into the evaluation process. This is important because of the new insights gained from varied perspectives and also because of the potential this gives for assuring the understanding, acceptance and utilization of findings. There are a variety of ways for accomplishing this.

The Texas rehabilitation program appears to have developed an evaluation process which includes features which should be a part of any evaluation cycle. The following model for an evaluation cycle draws heavily upon the cycle implemented in the Texas rehabilitation program.*

1. Each year, statements of problems and other program aspects needing evaluation are solicited from district administrators, supervisors, and counselors in the field, as well as from central office staff.

2. The full set of ideas are put before the State Director with a set of priorities recommended by the evaluation staff. Presumably, the recommended evaluation priorities are based on current overall program priorities as understood by the evaluation staff. The Director then formulates his own set of priorities and chooses the topics for study and evaluation that year. (These special topics are in addition to

---

*The source for information concerning the Texas evaluation process was Mr. W.F. Morehead, Assistant Director for Program Evaluation, Texas Rehabilitation Commission. The cycle described above is not completely identical with the Texas cycle, although the Texas agency's cycle clearly served as the inspiration and model.

routine program monitoring and ongoing evaluation, which are based primarily on management information using R-300 data.)

These two steps have several key advantages:

a. They assure that studies will only be done when the Director is interested in the results that could emerge from the study. This enhances utilization of the results, and assures that evaluation does not become a research enterprise far removed from program needs.

b. They provide field staff with a vehicle for getting operating problems of concern to the field before the Director. They also provide a vehicle to field staff for criticizing the usefulness of the evaluation unit's proposals for evaluation topics.

c. The steps compel the evaluation unit to make clear to the Director (and to itself) why it believes that proposed evaluation efforts will be useful. Thus, what the evaluation unit proposes to do is more likely to be prompted by program needs and priorities rather than academic research interest.

d. The steps bring to the evaluation new ideas from the field, while also helping the field staff understand the potential usefulness of evaluation to the agency program.

e. The evaluation program and priorities of an agency become updated annually.

Continuing on with the steps in this ideal evaluation process,

3. The evaluation unit <u>plans</u> and <u>conducts</u> the evaluation study with the <u>participation and cooperation of various field staff</u>.

4. When the study is completed, the evaluation unit writes a report. The unit is obligated in the report, however, not only to report the results of the study, but also to specify explicitly in writing

what the unit perceives to be the implications of the findings for the program and what specific program actions or decisions the evaluation unit would recommend based on the findings.

5. Before the report goes to the Director, however, it is sent down to the field (district administrator) and to the line divisions in the central office for review, comment, and rebuttal. The reviewers write down their reactions and propose -- if appropriate -- counter-conclusions and recommendations for action.

6. These reactions by field staff and the line divisions are sent along with the evaluation unit's report to the Director. Only at this point does the Director formally receive the evaluation findings. (The report apparently is not rewritten to reflect suggestions and criticisms by the reviewers prior to going to the Director. I gather that a rewrite may take place later.)

7. The Director reads the report and the reactions, reviews the situation, and makes whatever decisions he believes are appropriate. These decisions are then clearly communicated throughout the agency, and line division assumes responsibility for the implementation.

These steps strike us as having the following advantages:

a. They assure that the study design is appropriate to the problems. Field and line staff are usually very resistant to long questionnaires with vague questions which cannot be justified and carefully rationalized by the evaluation unit. (Indeed, they are resistent even when such questions can be justified.) Such operating staff also will resist study procedures which intrude on the time and energies of the operating staff. Since such intrusion is often necessary to find answers to the problems the evaluators are committed to explore, the involvement of such operating staff puts the burden on them to suggest alternate

designs. If such designs can be proposed, the evaluation unit should accept them. If such designs cannot be proposed, the evaluation unit has grounds for expecting fuller cooperation from the operating staff.

b. The involvement of operating staff helps them understand why the design is necessary, and thus serves to induce their commitment to the evaluation and to enhance the prospects of later utilization of the evaluation findings.

c. The steps assure that study findings will be linked to program actions. Another research report on the shelf does the agency no good. Knowing that they must recommend program actions, the evaluation unit will consider from the outset of the study -- rather than only at the conclusion -- what kinds of program action would be implied by alternative study findings. Such consideration of program implications usually changes the design of the study.

d. Obtaining reactions from field and line staff serves to correct any naive conclusions, misunderstanding of local operating procedures, and oversights which can creep into an evaluation unit's report. Since evaluation units are housed in the central office and are manned by people with more academic background in many cases, they can become fairly isolated from operating experience. (This isolation at the same time has the advantage of allowing them to think critically to ask new kinds of questions, and to look at the program objectively. Manning an evaluation unit only with former operating personnel has disadvantages too. The tension between evaluation and operating staff is healthy.)

e. Presenting the Director with the report and reactions by operating staff allows him to see all sides of the issues. It gives his decision greater legitimacy in the eyes of the agency staff. Finally, it tends to force the hand of the Director; some action or decision must

result from the study. Good administrators should tend to like this kind of decision-forcing approach. When they truly don't want to make a decision, they say so (and will probably kill the study at the outset, rather than expending all the time, money and energy of the agency on the study effort). Weak administrators, on the other hand, may find this approach threatening. It hampers their evasion of the decisions necessary to really attack and hopefully solve weaknesses in the program.

f. Pinpointing the responsibility for implementation in the line division helps assure that the Director's decree will indeed be carried. While to staff in agencies led by strong administrators, such a concern for the implementation of decisions may seem trivial, the concern will appear very legitimate in other agencies. The most typical situation in public agencies is that a Director's decrees are often implemented on paper or only half-implemented by district administrators and field staff, and sometimes even line divisions in a central office.

We clearly have idealized the Texas process in some ways. In particular, we are not well informed on who becomes involved in the design and conduct of evaluation studies, and how much. Also, we wonder if implementation proceeds as forthrightly as the above sequence suggests. The I.R.S. report perhaps should suggest to state agencies that:

1. They develop a mechanism for having representative counselors, supervisors, district administrators, and even consumers sit in on the selection of problems for proposed evaluation, on the design of such studies, and on the review of the findings of such studies. A task force of five or six people might be appointed to consult with the evaluation unit. Such consultation could occur by phone or mail, not by expensive gathering in the state capital. Indeed, since evaluation staff do more

traveling than field staff, the evaluation staff might consult with individual task force members while traveling for other reasons in his general geographical area. Also, most of the task force might be drawn from the area around the central office, minimizing the need for travel. The contributions these field staff and consumers might make to the realism of the unit's efforts are considerable. Even former counselors and administrators who move up into evaluation tend to lose their contact with current field thinking, even though they would not readily admit so. Note also that the rotation of appointments to this task force on a biannual basis would over time increase the number of people in the field (and in the consumer groups capable of lobbying the legislature) who are aware of the needs and problems of evaluation and of the rehabilitation program. By talking to their colleagues and friends, such individuals would improve the climate in the field which evaluation takes place.

This proposal is not that costly. A maximum of three to six days a year would be needed from each member of the task force for effective participation.

We would especially emphasize that disabled clients, as important actors in the rehabilitation system, could participate in the evaluation process via such a task force or advisory panel. Clients can be particularly helpful in identifying factors impeding program success from the client perspective. As for the selection of clients, it will not be of much benefit if clients who mimic the values of the agency or who are inarticulate are selected. If we really want feedback from clients in order to enrich our understanding and perceptions of problems, then clients should be selected who have ideas and suggestions and who are capable of articulating them. (NOTE: In California, several district

administrators are having client organizations talk to counselors about problems in service delivery from the client perspective. The initial experience has apparently been so rewarding that discussions have started about having such client evaluation and sessions with counselors become part of graduate school and in-service training programs.)

2. When the Director makes a decision following an evaluation study, he should not only specify who has authority for implementing the decision, but should also specify who has the responsibility for monitoring implementation and direct that a report be sent to him in six or twelve months on progress in implementation. Although this suggestion really reflects general principles of management, rather than of evaluation, such principles need to be restated as part of any discussion of evaluation. The challenge confronting state agencies is to make evaluation serve the purposes of management and decision-making. Many evaluation efforts currently produce only research documents. They look good on publication lists of the agency and the staff, but have little impact on the program and thus waste the taxpayer's money. This failure stems in part from the hesitancy of program directors to build in follow-through on the decisions which are made following an evaluation study.

B. Composition of the Evaluation Unit

An ideal evaluation staff should contain both specialists and generalists. There should be specialists who have graduate-level skills in experimental design methods and practical experience in the application of economic analysis and statistics. This will probably mean at least two persons since seldom does a single person combine all these skills. Additionally, because these people must apply their skill within an organizational context, it will be valuable for them also to have

personally, or to have available to them a person with knowledge of organizational theory, public administration, and what is often called the theory of planned change. Every member of an evaluation team, whether a technical specialist or a generalist, should ideally have the interpersonal skills needed to communicate with actors internal and external to the rehabilitation system. In addition to the above technical staff, every evaluation unit should have at least one member with strong field experience in rehabilitation counseling and administration. If one of the technical staff has such a background, excellent. If not, such a generalist should come from within the program and could perhaps be assigned annually on a rotating basis to the evaluation unit.

It is clear from the description above that an ideal evaluation unit would have from two to four individuals, depending on the availability of individuals combining the skills above. In most cases, three people will probably be sufficient. It should usually be possible to put together the following kind of team:

1. an individual with field experience in counseling and administration, who understands the nuts and bolts of program operations and who knows the theory of counseling and perhaps also management.

2. an economist, planner, or specialist in public or even business administration, who knows public economics, program analysis and P.P.B.S., cost-benefit analysis, organizational theory, the theory and politics of public administration, and some econometrics or social statistics. Such a person may often also be knowledgable about PERT, operations research, and similiar techniques, but these techniques should be given relatively lower priority in favor of broader skills in program analysis, administration, and policy planning.

3. someone who knows evaluative social research, some body of theory concerning social relations, experimental design and statistical analysis,(contingency tables and nonparametric statistics, analysis of variance and covariance, etc.). Most importantly however, such a person should be geared to thinking about policy and program decision, rather than pure research. Such a person should come from a range of disciplinary backgrounds: sociology, social psychology, public health, social policy planning, and even social welfare.

In an ideal team one member of the team (either #2 or #3) would come from a social policy background or from a social welfare background emphasizing policy and administration. Such a person would be knowledgable of the principles of planned organizational change. Increasingly, planning and social work schools are emphasizing such training, and more knowledgable individuals should increasingly become available to state agencies.

Each member of the above team will tend to look at the world and approach his work in very different ways. Each thus brings to the team some unique talents and perspectives. Probably, the P.P.B.S. specialist represents the newest and most valuable addition to such evaluation units as currently function in the rehabilitation state agencies. (In our own judgment, if the state agency cannot afford three people, we think individual #3 is the most dispensable.) If the state has money for four staff members, it would be most productive to add someone sharing the perspective of individual #1.

It is especially important that the head of the team be an individual who shares the personal confidence and trust of the agency Director. The possession of such a relationship far transcends in importance to the evaluation activity the possession of any technical skills.

Without such confidence on the part of the Director, no evaluation unit will be effective and useful -- no matter how competent and knowledgable the unit staff.

Evaluation staff should work closely with line divisions and not be seen as engaging in isolated activity not related to everyday program activity. As described above, some isolation is inevitable, but the unit should structure its activity to minimize the evaluation.

The skills of this evaluation staff should preferably be at a graduate level in training and should be bolstered by applied research experience in the use of statistics, economic analysis, and experimental method. This emphatically does not mean that Ph.D. status is required. Having taken only one or two courses, on the other hand, is usually not sufficient to meet the skill requirements. Given these requirements, evaluation personnel should not be drawn exclusively from among current rehabilitation personnel who have moved up through the ranks from an initial counseling position. Such a career ladder into evaluation has apparently been the dominant pattern in agencies to date, however. Thus, the need exists to:

1. revise graduate training in rehabilitation and social work
2. recruit others, from outside the field
3. teach, on the jobs, the skills needed for becoming a generalist.

C. Utilizing Outside Evaluators

In utilizing outside evaluators it is important that the agency assigns agency staff to work closely with the evaluator. This will increase the relevance of findings and the likelihood of utilization. The outside evaluators should have the specific skills mentioned above. Additionally, specific program knowledge would be useful on the part of

the outside evaluator. Where such knowledge is initially lacking, however, intensive discussions with agency staff and field visits can usually serve to inform the evaluator quickly.

Finally, if a program is sponsored jointly with another agency, evaluation might be more effective if it were sponsored jointly by the two agencies. This will increase the likelihood that any problems, such as in coordination, will be identified by both agencies and that effective joint action will result at the end of the evaluation effort.

# A BRIEF BIBLIOGRAPHY ON SOCIAL PROGRAM EVALUATION

The first group that follows consists of items for quickly getting an overview of issues and organizational and methodological problems relating to evaluation in general of social programs. The second group consists of selected key studies or overview essays relating specifically to rehabilitation programs.

## General References on Social Program Evaluation

American Institutes for Research, Evaluative Research: Strategies and Methods, (Pittsburgh: AIR, 1970) -- papers and summary of discussion at conference, good reading. Note especially the paper by David Hawkridge, "Designs for Evaluative Studies."

Francis G. Caro, "Approaches to Evaluative Research: A Review," in Louis Zurcher and Charles Bonjean (eds.), Planned Social Intervention: An Interdisciplinary Anthology, (Scranton: Chandler, 1970), pp. 403-421; also in Human Organization, 28 (Summer, 1969), 87-99.

Francis G. Caro, ed., Readings in Evaluation Research, (N.Y.: Russell Sage Foundation, 1971) -- a very fine collection of readings, including examples of applied evaluation research. Comparable in quality to the Weiss reader with more emphasis on methodology, but available only in hardcover. Contains the Caro summary article.

John W. Evans, "Evaluating Social Action Programs," in Zurcher and Bonjean; also in Social Science Quarterly, 50 (December, 1969).

Thomas K. Glennan, "Evaluating Federal Manpower Programs: Notes and Observations," RAND Corporation, Memorandum RM 5743-OEO (September, 1969) -- also in Weiss in abbreviated form.

Martin Rein, "Social Policy Analysis as the Interpretation of Beliefs," Journal of the American Institute of Planners, (September, 1971), 297-310.

Peter H. Rossi, "Practice, Method, and Theory in Evaluating Social Action Programs," in James L. Sundquist, ed., On Fighting Poverty: Perspectives from Experience, (N.Y.: Basic Books, 1969), pp. 217-234.

Edward Suchman, Evaluative Research (N.Y.: Russell Sage Foundation, 1967), minor "classic" of field, most often used text.

Tony Tripodi, Social Program Evaluation: Guidelines for Health, Education and Welfare Administrators, (Itasca, Illinois: F.E. Peacock Pub., 1971) -- a short and simple book with good overview of methods, fine introduction.

Carol Weiss, ed., Evaluating Action Programs (Boston: Allyn & Bacon, 1972) your BEST BUY; a comprehensive and uniformly excellent reader; contains shorter articles by Suchman, Glennan, Rossi, Wholey et al, Evans, etc., as well as many fine articles and papers not available elsewhere.

Joseph S. Wholey et al, Federal Evaluation Policy: Analyzing the Effect of Public Programs (Washington, D.C.: The Urban Institute, 1970) -- assessment of evaluation activities of OEO, HEW, HUD, LABOR.

Walter Williams, Social Policy Research and Analysis: The Experience in the Federal Social Agencies (N.Y.: Elsevier, 1971) -- similar to Urban Institute Study, with more examples, aimed at teaching evaluation.

References Specific to Rehabilitation

Larry Biscamp, Charles Cole, Judy Taylor, and Herbert Willsmore, "A Client Evaluation of Rehabilitation Counselor Training Programs," Institute of Urban and Regional Development, University of California, Berkeley, Working Paper No. 176/RS006.

Julius Cohen, Irene Butter, Stanley Deline, and Ronald Nutter, eds., Benefit-Cost Analysis for Mental Retardation Programs: Theoretical Considerations and a Model for Application (Ann Arbor: Institute for the Study of Mental Retardation and Related Disabilities, University of Michigan, 1971).

Frederick Collignon, Adam Zawada, Barbara Thompson, and Joel Markowitz, "Guidelines and Criteria for Evaluating Vocational Rehabilitation Programs' Institute of Urban and Regional Development, University of California, Berkeley, Working Paper No. 173/RS003.

Ronald W. Conley, "A Benefit-Cost Analysis of the Vocational Rehabilitation Program," Journal of Human Resources, IV, No. 2 (Spring, 1969), pp. 226-252.

Ronald W. Conley, The Economics of Mental Retardation, (Baltimore: Johns Hopkins Press, 1972).

Ronald W. Conley, The Economics of Vocational Rehabilitation, (Baltimore: Johns Hopkins Press, 1965).

Charles Grigg, Alphonse Holtman, and Patricia Martin, Vocational Rehabilitation of Disabled Public Assistance Clients: An Evaluation of Fourteen Research and Demonstration Projects. Institute for Social Research, No. 8, (Tallahassee, Florida: The Florida State University, 1969).

Elizabeth A. Heferin and Alfred H. Katz, "Issues and Orientations in the Evaluation of Rehabilitation Programs: A Review Article," Rehabilitation Literature, Vol. 32, Part I (March, 1971), pp. 66-73, and Part II (April, 1971), 98-106.

Michigan Department of Education, Division of Vocational Rehabilitation. "The Vocational Status of Michigan Rehabilitants of Fiscal Year 1969 Two Years After Case Closure, The Results of a Follow-Up Study and Benefit/Cost Analysis Conducted by the Program Analysis, Planning, and Development Section of the Division of Vocational Rehabilitation." Michigan Department of Education, Division of Vocational Rehabilitation, February, 1971.

Saad Z. Nagi, Disability and Rehabilitation, Legal, Clinical and Self-Concepts and Measurement. (Ohio State University Press, 1969).

Donald Schon, "The Blindness System," The Public Interest, 18 (Winter, 1971) pp. 25-38.

Marvin B. Sussman, ed., Sociology and Rehabilitation (American Sociological Association, 1966). Note in particular the essay by Edward Suchman, "A Model for Research and Evaluation on Rehabilitation."

Pat F. Sutherland, "Program Evaluation in Social Action Programs, with Glossary," State of Texas Commission for the Blind, 1971.

George N. Wright, Kenneth W. Reagles, and Alfred J. Butler, The Wood County Project, An Expanded Program of Vocational Rehabilitation (Madison, Wisconsin: The University of Wisconsin Rehabilitation Research Institute, 1969).