

Originally Processed With FOIA(s):

S

FOIA Number:

S

FOIA MARKER

**This is not a textual record. This is used as an
administrative marker by the George Bush Presidential
Library Staff.**

Record Group/Collection: Donated Historical Materials
Collection/Office of Origin: Frieden, Lex, Collection
Series: Printed Materials
Subseries: Manuals

OA/ID Number: 52104
Folder ID Number: 52104-010

Folder Title:
"A Resource Handbook for Vocational Rehabilitation" [1974]

Stack:

Row:

Section:

Shelf:

Position:

PROGRAM EVALUATION

A RESOURCE HANDBOOK FOR
VOCATIONAL REHABILITATION



PROGRAM EVALUATION: A Resource Handbook For Vocational Rehabilitation

Eleanor Carol Bennett

Resource Analyst

Research Utilization Laboratory

ICD Rehabilitation and Research Center

Marvin Weisinger

Associate Project Supervisor

Research Utilization Laboratory

ICD Rehabilitation and Research Center

**ICD Rehabilitation and Research Center
New York, N.Y. 10010**

Copyright © 1974 ICD Rehabilitation and Research Center except for portions copyrighted by others, permission for the reprinting of which herein has been obtained from the respective copyright owners.

Portions of this publication, except for those portions copyrighted by others, may be reprinted provided permission to do so is obtained in writing from the ICD Rehabilitation and Research Center or from the Social and Rehabilitation Service, U.S. Department of Health, Education, and Welfare with whose support the project resulting in this publication was conducted.

The development of this publication was supported, in part, by Research Grant No. SRS 22-P-55060/2 from the Division of Research and Demonstration Grants, Social and Rehabilitation Service, Department of Health, Education and Welfare, Washington, D.C. 20201.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
PREFACE	vi
I. PROGRAM ASSESSMENT	1
OBJECTIVITY	2
A PERSPECTIVE	5
A SUGGESTED MODEL	7
Classes of Measures	7
Program Level	9
A Word on Indices	11
Some Additional Considerations	11
SUMMARY	15
II. PROGRAM EVALUATION DESIGN	17
STATEMENT OF PROGRAM EVALUATION PROBLEM	18
Operationalization	19
SPECIFICATION OF METHODOLOGY	21
Approaches to Evaluation	21
Research Approach	22
Systems Approach	24
Reliability, Bias, Validity	27
Selection of the Sample	28
COLLECTION OF DATA	29
ANALYSIS OF DATA	32
Levels of Statistical Measures	32
Data Analysis Techniques	33
Research Approach	33
Systems Approach	40
INTERPRETATION AND CONCLUSIONS	43
III. THE NINTH PANACEA: PROGRAM EVALUATION	45
By Robert A. Walker	

IV. COMPENDIUM OF MEASURES	61
PROGRAM	63
Effort	63
Effectiveness	64
Efficiency	70
Quality	71
SERVICE DELIVERY	72
Counselor Performance	73
Evaluator, Job Trainer, Placement Officer	75
PROGRAM MANAGEMENT	76
Administrative Functions	77
Community Relations	78
Staffing	81
Inservice Training	82
Fiscal Administration	84
WEIGHTED CLOSURES	85
CASE FLOW (Monitoring)	86
V. EXAMPLES OF PROGRAM EVALUATIONS	89
PLANS	90
RESEARCH APPROACH	95
SYSTEMS APPROACH	99
COST BENEFIT ANALYSIS	104
OTHER FIELD EXAMPLES	109
GLOSSARY	111
READING LIST	117
REFERENCE LIST	139
APPENDIX I. Other Classes of Measures Used in Program Evaluation	149
APPENDIX II. Standards	150

ACKNOWLEDGEMENTS

Compilation of this handbook reflects the support and suggestions of many people. Special acknowledgment must go to the staff of the Research Utilization Laboratory whose aid and encouragement was instrumental in its completion—Dr. Isabel Robinault, Project Supervisor and Susan Ahrenhold, Information Specialist. Dr. Salvatore G. DiMichael, Project Director and Director of ICD Rehabilitation and Research Center provided valuable assistance to this effort by placing the Center's resources at our disposal. Helen Stonehill, Center Librarian, generously supplied us with much of the esoteric literature in the field of program evaluation. In addition, we must express our great appreciation for the fortitude and patience of our secretarial staff, Helen Carr and Christine Hicks, who typed numerous drafts of this document. Finally, we wish to thank the many program evaluation professionals whose substantial consultant effort is critical to enhancing the Handbook's value for the field. Any errors in the interpretations and compilations based on the consultant's input for the program evaluation examples and compedium of measures are ours alone. We particularly would like to acknowledge the efforts of the following:

Edward C. Baumheier, Ph.D.
Janet M. Derr
Lalitha Raichur
Social Work Research Institute
University of Denver
Denver, Colorado 80210

James Bitter, Ph.D.
Regional Rehabilitation Research Institute
University of Northern Colorado
Greeley, Colorado 80631

Gabriel Cifor
Robert Struthers, Ph.D.
Department of Education/
Vocational Rehabilitation Services
Lansing, Michigan 48904

Larry Dickerson, Ph.D.
Research & Training Center
University of Wisconsin-Stout
Menomonie, Wisconsin 54751

William Hills, Ph.D.
Barbara Chesney, Ph.D.
Regional Rehabilitation Research Institute
The University of Oklahoma
Norman, Oklahoma 73069

Adrian J. Marinelli
Research Utilization Specialist
New Jersey Rehabilitation Commission
Trenton, New Jersey 08625

Leonard Miller, Ph.D.
Division of Counselor Education
College of Education
University of Iowa
Iowa City, Iowa 52242

John Muthard, Ph.D.
Regional Rehabilitation Research Institute
University of Florida
Gainesville, Florida 32601

Robert Philbeck
Assistant Director for Planning & Evaluation
Division of Vocational Rehabilitation
North Carolina State Department of Human
Resources
Raleigh, North Carolina 27602

Sidney Stein
Program Evaluation
Division of Vocational Rehabilitation
Texas Rehabilitation Commission
Austin, Texas 78705

Adam Zawada
Program Planning and Research
Division of Vocational Rehabilitation
Tallahassee, Florida 32304

PREFACE

This handbook on program evaluation is a compilation of approaches, methods, and techniques culled from over 250 sources. It is an attempt to bring together a large array of information and organize it into a unified conceptual framework.

The evaluator, whether novice or experienced, often faces a bewildering mountain of information when he attempts to examine the literature of program evaluation. There seems to be no one source where information on indices can be obtained. Indeed, from publication to publication, there seems to be little or no agreement even on definition of terms. Therefore, one of the major efforts of this manual was to review the diverse literature and opinion on program evaluation, to organize it, and reduce it to a useable form. However, it should be clearly understood that while the scope of the Handbook is comprehensive, we have not attempted to write the "magnum opus." *Program Evaluation: A Resource Handbook* is in no way a substitute for a text book, background training or experience. Some of the defined approaches and techniques require a good deal of technical sophistication to be properly applied, interpreted, and used in making program decisions for change. Some techniques and approaches are relatively simple to apply and this is indicated. Most important, we have tried to be cautionary where appropriate.

The contents of this manual are directed to three audiences. One audience is comprised of the new evaluators. We assume they have been exposed to a course in statistics and a course in research methods, have read several program evaluations and are now wondering how to put it all together. The second audience consists of experienced evaluators who might be looking for some specific measures or techniques to solve an immediate problem. For the former, we hope this manual will serve as a major adjunct to his training and give him some important leads, cues and cautions on how to organize and proceed with a program evaluation. For the latter, the annotated list of indices in Chapter IV and the Bibliography and Reading List could be most helpful. The third audience consists of administrators. For them, we hope this handbook will serve as an aid to communication with program evaluation staff.

Since program evaluation is in its relative infancy in some of the social services and since our review of the literature indicated little consensus among authors, we expect certain honest differences of opinion regarding the material contained in this handbook. We hope this document will be read, discussed and, perhaps, even argued over.

New York
June, 1974

Eleanor Carol Bennett
Marvin Weisinger

**PROGRAM EVALUATION:
A Resource Handbook For
Vocational Rehabilitation**

I PROGRAM ASSESSMENT

Today, program assessment is a key concern of management—both governmental and private. Controversy surrounds it and published contradictions are rampant. Some of the reasons for this controversy and lack of understanding is the relatively recent use of program assessment in the management processes of human services where its purpose may seem antithetical to these services. In addition, the personnel involved in assessing programs and those utilizing the results of such assessments often have conflicting philosophies and differing goals. Consequently, the information-decision-making value of program assessment can get lost or may be interpreted as a spectre capable of damaging worthy programs, productive jobs and creditable reputations.

Part of this difficulty lies in the differing backgrounds (and thus differing philosophies) of the evaluator and the practitioner-administrator. In many cases, the administrator who received his primary training as a practitioner in the helping services can be quite astute in systematically analyzing the complex problems of clients and recommending appropriate actions. Yet, when he is with systematically analyzing the complex problems of a program and recommending actions, he is somewhat at a loss. In addition, the training of a

practitioner emphasizes the solution of client problems with all due speed, without too much regard for economy.

On the other hand, the evaluator's background stresses precision and control. He is often at a loss to understand why the boundaries for acceptable performance for program activities are set so wide.

Communication between evaluator and administrator is often at cross purposes, even though they may be discussing the same things, they use different words. One purpose of this handbook is to provide an administrator with a conceptual framework which will aid him to effectively communicate with evaluation staff and to aid him in analyzing his program with the same acumen he analyzes client problems.

It must be clearly understood that the final outcome of any evaluation process is *change*. Evaluation basically is a systematic and objective method of looking at where a program presently stands and estimating how a program can do better. Change is an integral part of this examination—change to serve the clients better and change in program economy. Only through an orderly process of documented program evaluation can justifiable alternatives for action be developed for the evolution of more efficacious and economical programs.

OBJECTIVITY

What sort of information should be used to document current program activity and to justify future change? The answer that comes quickly to mind is, "objective information". Two questions can then be asked, "How objective can an assessment be?" and "How does one objectively assess a program?"

A program assessment can be conducted on one of two major levels of objectivity. The least objective, Level I, is the opinion of experts. That is, it is an assessment based upon the review of project documents or based on an onsite visit by a person recognized as "expert" in the field. The review may be superficial or in depth, but it relies on the previous experience of the reviewer for the ultimate statement of program worth. It is quite conceivable that two experts, reviewing the same program, could present two very different sets of findings. Therefore, choice of this level of objectivity should be avoided whenever possible.

The more objective approach, Level II, is the systematic collection and analysis of hard data. If this approach is followed, it should produce findings that are reproducible no matter who undertakes the assessment.

Who administers the program assessment depends on the organization of the agency, the finances available for the assessment and the source of the request for assessment. Usually, there are three categories of people assessing programs: consultants or consulting firms, panels of knowledgeable people and departments within the organization.

Exactly who administers the appraisal many times influences the impact of the assessment. For instance, an expert visiting the program on a site visit, or an outside researcher collecting "relevant" data to generate "meaningful" comparisons can be open to attack on such dubious grounds as "lack of familiarity with local conditions" or the researcher's "Ivory Tower" approach. On the other hand, an "in-house" assessment may be accused of a "white wash" or bias. There is no simple solution to the selection of the evaluator. However, the use of a rigorous objective approach that is replicable can usually be defended against such attacks no matter who administers the assessment.

The answer to the second question "How does one objectively assess a program?" is dependent on a number of factors: the type of information available, the precision required from the conclusions, and the funds available to undertake this assessment. There are three basic means of program assessment, that is, a program may be monitored, evaluated, or be the subject of a special study.

Monitoring is ongoing feedback of information about professional performance and program activities for the purpose of controlling those activities. This type of assessment serves as an indicator of disruption in program activities, but does not necessarily define the nature of the difficulty. No value judgement is offered concerning the "goodness" or "badness" of the program, although application of externally or internally derived standards may offer some kind of statement concerning professional activities. Examples of usual monitoring methods are caseflow information and case review. In some instances, ongoing outcome measures are examined to indicate disruption of activities.

Program Evaluation usually examines annual program information and can offer a value judgement on an entire program's worth. It can identify causes of problems or reasons for success. Usually this judgement is based on program outcomes, or the outcomes of program components.

Special studies offer a judgement from the examination of one facet of a program. That facet may be universal, in that it pervades many portions of the program, or it may be highly circumscribed. The final impact of that assessment depends upon what aspect of the program has been examined. For instance, to examine "the effectiveness of counselor performance when serving the culturally disadvantaged" is a pervasive problem encompassing attitudes and program policy. On the other hand, to examine "the effectiveness of white counselor performance in dealing with aged, black alcoholics" is specific to the administration of one program and, as such, is so circumscribed as to be of doubtful utility.

Figure I illustrates some of the benefits and limitations of the combinations of level of objectivity and level of assessment. Since the methods, dynamics and value of monitoring are well understood as necessary mechanisms for administrative control, our discussion of this approach is limited to a presentation of indices for monitoring in Chapter IV.

It can be seen from the above discussion that objectivity is partially a matter of the degree of refinement of information from which judgements are to be made. The remainder of this monograph is addressed to that process which is based on the analysis of hard data and is used to evaluate programs. This technique, *program evaluation*, is just now emerging as a program-planning and decision-making tool. "*Special Studies*" may be considered as a limited or circumscribed case of the broader "Program Evaluation". Therefore, any of the following discussions which refer to Program Evaluation may be considered as perfectly applicable to the topic of Special Studies.

FIGURE I
BENEFITS AND LIMITATIONS OF
PROGRAM ASSESSMENTS

Level of Objectivity	Level of Assessment		
	Monitoring	Special Studies	Evaluation
I. Expert's Opinion from A. Site Visit B. Annual Report	Benefits	Benefits	Benefits
	Inexpensive	Inexpensive	Inexpensive
	Not time consuming	Not time consuming	Not time consuming
	May identify problems	May be perceptive	May be perceptive
	Limitations	Limitations	Limitations
	Not replicable	Not replicable	Not replicable
	May not identify problems	May be imprecise	May be imprecise
	May not identify cause	Lack of adequate firm data	Lack of adequate firm data
II. Analysis of Hard Data	Benefits	Benefits	Benefits
	Ongoing	Precise	Precise
	Identifies problems	Replicable	Replicable
	Replicable	Identifies cause	Identifies cause
	Limitations	Limitations	Limitations
	Does not identify cause	Expensive	Expensive
	More expensive than Level I	May be time consuming	Time consuming
		Limited in scope	

A PERSPECTIVE

In reviewing the program evaluation literature, several problems confront the reader. A great deal of literature is directed to extolling program evaluation's virtues as an aid in management, and an equally large collection decries the damage it can and does do, if perspective and methodology are not sound. A smaller body of representative literature discusses technique and methodology. However, these methodology articles are written from what appears to be many conflicting points of view. Actually, close examination reveals only two or three points of view.

The first point of view is that of administrators who have learned to analyze their programs by using several different models. Probably the most popular model speaks of *input analysis*, *process evaluation* and *outcome evaluation*. A second model refers to levels of program responsibility or control and mentions *the institution* (program), *managerial functions*, and the *technical core* (service delivery). (95)

An administrator, when considering his program, may view it as a system to be controlled and regulated. He then speaks of specific parts such as inputs, process, outputs, and some sort of a "feedback loop".* The analyses of these parts may be defined as follows:

Input Analysis—the study of the resources (men, money and materials) used by the program to accomplish its goal.

Process Evaluation—the study of a combination of program operations and techniques used by a program to accomplish its goal. Criteria usually are addressed to questions of conformity to, and expectations of the respective professions involved. These criteria are usually called *Quality*. Many times the statistics used to monitor programs are composed of these types of data.

Outcome Evaluation—the study of the results of a program's activities. Usually, the criterion used is the degree to which programs have attained goals. Outcome evaluation is frequently obtained through follow-up studies.

The second model relates to specific levels of program responsibility (responsibility model)—the institutional level, managerial level and technical level. This breakdown may be considered as a reduction to *program level*, that is:

*Program monitoring and evaluation are part of the feedback loop.

Institution—(a program) is supported by a larger social system which legitimatizes the program's goals. Stated another way, we can say society expects a vocational rehabilitation program to assist the disabled to become meaningfully employed.

Technical—(service delivery) is a sub-organization or level of program. Its function is to provide the services to clients and so accomplish the program's goal.

Managerial—(program management) is another sub-organization or level of program. Its function is to mediate between services, and between services and clients. In addition, it procures resources and aids in assuring appropriate outputs.

A second point of view is that of the evaluator who usually addresses his perspective to methodology. This perspective can be divided into two camps—the researcher who speaks of comparisons and goal attainment, and the systems analyst-operations researcher who speaks of effectiveness and survival models. (34) For the purposes of practical illustration, the popular goal-attainment model may be considered as an *inductive* approach, while the systems models usually are *deductive* in nature. Research constructs hypothesis about the way programs work, and the systems approach builds models describing program structure, functions and goals. Below are our definition and distinctions of the two approaches:

Research Approaches seek to evaluate programs through making certain comparisons between variables. The basic question is, has the program caused a change in a variable? For example, has the introduction of a PA/VR program caused more PA recipients to leave the welfare rolls? These comparisons may be: before vs. after introduction of a program, between programs operating in differing localities, between clients in a special program and clients who did not have the benefit of the special program, between client behavior before entering the program and after leaving the program, etc. Of course, certain steps must be taken to assure comparability of the data (See Chapter II).

A special case of the research approach is the *Goal Attainment Model*. In this case, in order to determine program effectiveness, the broad objectives of the program are translated into measurable criteria and a comparison is made between these criteria and actual program performance, i.e., has the program attained its goal? *Standard Setting* or *Comparison of Planned vs. Actual Performance* are special cases of the Goal Attainment Model in which the goals are set by some formula or by professional judgement and

are not necessarily one-to-one reflections of total program objectives. (See Chapter III.)

Systems Approaches—seek to evaluate programs by describing the organizational structure of a program, the stated functions of the program, and the process by which that organizational structure with those functions accomplishes the program's objective. That is, the analytic process tries to build a model of the program, uses the model to stimulate the program, and then offers more efficient and/or effective alternatives of action (optimization). An enormous amount of information is needed for this type of evaluation, together with a great deal of data processing effort. (See Chapter II.)

A SUGGESTED MODEL

"The purpose of the evaluation process is to provide policymakers with the basic data necessary for them to make decisions wisely. Impact evaluations of programs should provide five essential sets of information. First, they should provide all of the data necessary to determine if a particular program should be continued. Second, they should determine which of the alternative programs achieve greatest gains for a given cost. Third, evaluations should present information on the components of each program and the mixes of components which are most effective for a given expenditure so that maximum operating efficiency can be achieved. Fourth, evaluation should provide the first three types of information for persons with different characteristics so that a decision-maker may determine which individuals are best served by each program. Finally, in the course of evaluating existing programs, data should be gathered which will suggest new methods to attack problems. To date, no evaluation of programs has provided all this information."(11)

In short, the primary goal of program evaluation is to ask "Which programs are worthwhile?" or "Which programs work best?" and then state "Why they are worthwhile" or "Why they work best." These findings may produce information for program decision-making that may result in the implementation of a change.

Suggested below is a model that approaches some of the issues for the systematic evaluation of programs.

Classes of Measures

Basically, there are five classes of measures or criteria that are used to estimate various aspects of program worthiness. The first class of measures enumerates the resources used by the program (effort). The second class ex-

amines the results of program service in terms of broad social change (impact). The third class attempts to estimate specific client outcomes or changes (effectiveness). The purpose of the fourth class is to determine the economics of program operation in comparison with program accomplishment (efficiency). The final class of measures, *Quality*, has come to mean different things to different people. One accepted global definition is that quality is a statement of program "goodness." A second definition used equates quality with effectiveness. That is, it uses process measures (see appendix I) to represent program effectiveness. The perspective taken in this handbook is that Quality estimates the conformance of program activity and professional activity to accepted standards of the day. It is one reading of ongoing program process. These five measures are derived from the administrator's system breakdown of input, process, and output, where effort represents input, quality is equivalent to process, impact and effectiveness represents output, and efficiency is an input-output ratio. More specifically:

Effort is a statement of fact about the "amount and kinds of program activities used to reach program objectives." It refers to staff time, activity and commitment, and to the allocation and use of material resources (funds, space, equipment, etc.). In addition, it takes into account ancillary resources used (outside consultation, media, public relations, etc.). It answers the question "What did I do?" (94,97)

Impact is a statement based on how the program was able to affect a broad social change together with the utility of this program technique. Usually, cost/benefit or cost/effectiveness is used for this type of evaluation. *(11)

Effectiveness is a statement of fact about how well the objectives of the program were met, or a statement of fact about how well the program components functioned to meet the program objectives or the program components' objectives. Effectiveness measures answer the question "Has Vocational Rehabilitation program accomplished its objective?" or "Have clients, who are rehabilitated by Vocational Rehabilitation, gotten jobs?" In other words, "How has the program changed clients behavior or performance?" (27,41,94,97)

*Impact is the same as effectiveness, but on a much broader scale; that is, it can be thought of as a long-range socially pervasive change, whereas, effectiveness is usually thought of as simply attaining goals or a circumscribed change. In many cases, impact criteria are very difficult to establish and data very costly to collect. Therefore, this monograph only considers effectiveness.

Efficiency is a rate that is produced when *effort* (inputs, staff time, activities, salaries, physical facilities and purchased service) expended by a program is compared to the program's *effectiveness* (output). It answers the question "Can the same program results (output) be achieved by either reducing the effort expended or by choosing other less costly alternatives of action?"(26,49,97)

Quality is a judgemental statement made about professional competence, acceptability of services to the client and the setting of minimum performance standards by legislative and regulatory bodies. It answers the question "How well did you do what you did?" or "How well did you use the knowledge you had?"(2,29,97)

Program Level

Having established and defined proxy measures for the system elements (input, process, outcome) the "responsibility model" (program, program management and service delivery), was selected as a suitable outline for analysis. The goals of each organizational element are defined below.

Program—goals are to meet society's expectation of a Vocational Rehabilitation organization.

Program Management—goals are to secure and organize inputs for the system (funds, staff, facilities, referrals) and then assure appropriate outputs (closures).

Service Delivery—(technical component)—goals are to perform the process of rehabilitating the disabled.

Figure II visually represents the conceptual outline for a program evaluation model that will include input analysis, process analysis, and outcome analysis (goal attainment) or measures of effort, effectiveness, efficiency and quality broken down by program level.

FIGURE II
PROPOSED MODEL FOR
PROGRAM EVALUATION IN
VOCATIONAL REHABILITATION

	MEASURE			
PROGRAM LEVEL	EFFORT	EFFECTIVENESS	EFFICIENCY	QUALITY
Total Program				
Program Management				
Service Delivery				

Using this matrix, it can be seen that any program level may be assessed by any class of measures. For example, a program evaluation may measure the effectiveness and efficiency of the entire program and the efficiency of program management, service delivery, etc. It is this procedure of measuring the program levels with multiple measures which will provide the most powerful (therefore the most defensible) program evaluation.

Many of the other terms which are so frequently discussed when referring to program evaluation, are specific methodologies or mathematical techniques used in estimating program effectiveness or efficiency. For instance, cost benefit is a measure of impact, while case review is a way to judge the quality of counselor performance, but more importantly, it is a way to gather data for studying program effectiveness.

A Word on Indices

Indices are counts, rates, ratios, or scales used to represent a class of measures. For example, a popular index is "the number of clients who gained jobs." This is used as an indication of program effectiveness. The relatedness of job placement to vocational training may be used to indicate the quality of counselor performance. There are many similar administrative yardsticks that are currently maintained and comprise a compendium of program evaluation measures. Figure III, is an illustration matching some of these popular indices with the specific measures and program levels.

Indices also provide elements necessary for the development of standards. For instance, a ratio made between annual accepted cases and cases closed status 26 may be used as a standard for monitoring programs. However, a second element necessary in the development of a standard is a judgement of the usefulness of that ratio as a standard. This judgement is usually arrived at by one of several different methods: the use of expert opinion, normative standards, the use of empirical evidence from outstanding research efforts, or the use of ratios from outstanding state programs.

Some Additional Considerations

Development of Evaluation Criteria from Program Objectives. A program may be considered worthwhile when it meets two basic requirements; the objective (i.e., goal or purpose) of the program is reasonable, and the program reaches this goal.

In most cases, consideration is not given to the *reasonableness* of program objectives. Therefore, no consideration is given to the reasonableness of the criteria developed from program objectives to measure program effectiveness. That is, if a program objective is to move the handicapped into gainful employment, one possible effectiveness criterion is the number of clients that are employed some time after completion of the program. In many instances, however, program objectives cannot be so easily translated into evaluation criteria because these objectives are not really reasonable. Some avowed program objectives may consist more of rhetoric than potential, because they may have been established only in response to political considerations. In other cases, the statement of program purpose is so vague and diffuse that development of measurable criteria that reflect these objectives becomes exceedingly difficult if not impossible. Further, to try to use evaluation criteria that attempt to reflect overstated or vague program objectives may result in an evaluation that "proves" the program a failure. It is only with ingenuity that the evaluator can come up with reasonable criteria against which the program can be measured. In addition,

FIGURE III
EXAMPLES OF INDICES USED IN PROGRAM EVALUATION

PROGRAM LEVEL	MEASURES			
	Effort	Effectiveness	Efficiency	Quality
1. Program	Annual No. of Clients	$\frac{\text{Annual No. of Status 26}}{\text{Annual No. Status 08,26,28,30}}$	$\frac{\text{Annual VR Budget}}{\text{Annual No. Status 26}}$	$\frac{\text{Average Weighted Closure Status}}{\text{Adherence to Regulations}}$
2. Service Delivery	Annual Cost per Service	Annual No. Status 26s	$\frac{\text{Total Service Cost}}{\text{Annual No. of Status 26}}$	Adherence to Regulations
a. Counselor Performance	No. of Contact Hours per Client	$\frac{\text{Annual No. Status 26}}{\text{No. of Counselors}}$	$\frac{\text{No. Contact Hours per Counselor}}{\text{Annual No. Status 26}}$	Rehabilitation Counselor Rating Scale
3. Program Management	Program Management Budget	Profile of Organizational Characteristics	$\frac{\text{Management Budget}}{\text{Annual VR Budget}}$	
a. Administrative Function		Procurement of Satisfactory Sources for Input and Output From the Program		
b. Community Relations	Cost of Community Relations Functions	$\frac{\text{Annual No. Acceptances}}{\text{Annual No. Referrals}}$ (See Chapter IV)	$\frac{\text{Annual No. Acceptances}}{\text{Annual No. Referrals}}$	$\frac{\text{Cost of Community Relations}}{\text{Nature of Inter-agency Agreements}}$

PROGRAM LEVEL	MEASURES			
	Effort	Effectiveness	Efficiency	Quality
c. Staffing	Annual \$ Spent on Recruiting	No. of Staff Hired in Job Description	$\frac{\$ \text{ Spent on Recruitment}}{\text{No. of Special Staff Hired}}$	Experience of Staff
d. Inservice Training	No. of New Techniques Presented to Staff	$\frac{\text{No. Who Use Techniques}}{\text{No. of Staff Who Could Use Techniques}}$	$\frac{\$ \text{ Spent on Training}}{\text{No. 26 Closures After Training} - \text{No. 26 Closures Before Training}}$	Experience of Training Staff
e. Fiscal Ad- ministration	Knowledge of No. of Sources of Funds Avail- able	$\frac{\text{No. of Sources of Funds Received}}{\text{No. of Available Sources}}$	$\frac{\text{Annual VR Budget}}{\text{Total Cost of Fiscal Administration}}$	Adherence to Regulations

he must use his abilities to convince decision makers that these criteria truly reflect program objectives, and that the resultant findings truly reflect program accomplishment.

The second factor which should be taken into consideration is the *time frame* of the program and the *environmental pressures* that impinge upon the program during the period being considered. For instance, a day care program was established in a district to assist AFDC mothers to work. Some time after the initiation of the program, this district's job market became depressed. The objectives of the program were, of course, not met—not because day care does not work, but because the objective is presently not reasonable. It might be noted in passing, that there are instances when a program objective is not reasonable at the time a program is planned. In the above example, if the day care program was planned when the job market was already depressed, the objective is unreasonable and the program is doomed to failure in advance.

The Environment. In too many cases, programs are evaluated as if they were operating in a vacuum. In the illustration above, it was a depressed job market that was unexpected at the time the program was planned that caused the ultimate “failure” of the program. This is only one environmental influence that can impinge on a program and alter measures of final outcome. It is the astute evaluator who will look at the environment in which a program operates for such variables as other programs competing for clients, job market, social acceptance of the program by the public and decision-makers, etc. No program functions outside of its environment, and careful consideration of these environmental factors is important in any program evaluation.

Goals and Time Frames. Many times the evaluation literature speaks of immediate, intermediate and long-term evaluating goals. Usually, these time-frame types of goals are used in discussions concerning evaluations of innovative or demonstration programs. In most instances, immediate goals are effort measures and question whether the program has invested the necessary resources to make a program work. Intermediate goals are usually quality measures, and they question whether the program staff have followed the clinical techniques set forth. Long term goals are usually effectiveness measures and they question whether the program outcome was satisfactory, given that the necessary resources were accumulated and the prescribed techniques were used.

The “Follow Up” Study. An entire program may be evaluated by following up clients, with given variables or indices (had job at closure), showing a client's behavior change. These follow-up measure always show outcome and therefore are effectiveness measures. In this handbook we will speak of “Effectiveness Studies” which refer to the type of measures used, rather than follow-up studies which speak of a point in time in a client's case history. In short, both Effectiveness Studies, Outcome Studies, and Follow-Up Studies are one and the same thing!

Evaluation of Sequential Activities. It is not appropriate to evaluate the effectiveness of any single step when that step is part of a sequential set of activities. That is, it is not feasible to evaluate "placement" effectiveness when the index used is dependent on the performance of the client "evaluation" and "training" steps. For instance, if the evaluation staff knows the training department wants to maintain classes of 20 clients and it recommends fifteen suitable clients and five less suitable clients, training may then "train" all twenty, but placement still may only be able to place the fifteen "best" clients. Is placement then ineffective? The exception to this rule is the referral step and the final outcome step. Referral is not preceded by any program activity and outcome reflects the total program accomplishment.

SUMMARY

This manual takes the position that the purpose of program evaluation is to:

- determine which are the best programs and why;
- suggest ways to strengthen existing programs;
- suggest new ways to solve problems.

This is best achieved by the analysis of hard data, using measures of effort, effectiveness, efficiency and quality. Hard data measures are rates, ratios and index numbers which have been extracted from counts, costs, time, rating scales, psychological test results, etc. The most powerful evaluations are those in which both the entire program and program components are examined.

Because of their ability to evaluate all aspects of a program, the most appropriate approaches are the Research Approach, which examines program objectives, and the Systems Approach, which examines the program structure and functions used to reach program objectives.

Further, follow-up studies which refer to a point-in-time in a client's case history are actually measuring program effectiveness. Therefore, follow-up studies and effectiveness studies are one and the same thing. The term describing the measure (effectiveness) is preferred over the term describing the point in time (follow-up).

II PROGRAM EVALUATION DESIGN

Basic to any scientific study is the requirement that the research is conducted in as objective a manner as possible. Ideally, in science, experiments must be conducted in such a way that any scientists examining the procedures used would agree that the researcher has conducted a reasonable and fair (unbiased) test of the phenomenon under study. They would also agree that any samples (selection process), which he used were adequate representations of all the material available and that he has carefully isolated (controlled) and compared what he wanted to study. In addition, the phenomenon under consideration in the research has been carefully identified and defined so that there can be no question in the mind of any scientist reading the study of exactly what phenomenon was examined and what was done. Finally, if a second researcher repeated the exact experiment, he should obtain the same results (reliability).

While this illustration is couched in terms relating to research in the basic sciences, it must be clearly understood that the scientific method has nothing to do with the test tubes or microscopes. Rather, it is a way of thinking—a way of objectively analyzing a problem and coming to conclusions. For instance, in

program evaluation, the phenomenon under study is the "program" and the area of examination is how effectively and efficiently that program met its objectives, whether the objectives were reasonable (worthwhile), and what parts (structures) of the program contributed to its success or failure.

The specific steps that a new evaluator should consider in doing a program evaluation are presented below. A discussion of each follows directly.

- a statement of the program evaluation problem
- specification of methodology
- collection of the data
- analysis of the data
- interpretation of the data and conclusions

STATEMENT OF PROGRAM EVALUATION PROBLEM

The basic goal of a program evaluation is to estimate the worthiness of a program. But "worthiness" is much too diffuse and ill-defined a concept to be of much practical use for developing a program-evaluation plan. Instead, development of a statement of a program evaluation problem (much like the "statement of the problem" in a formal research study) is of great importance in clarifying and delineating what is going to be done in the evaluation.

Generally then, a statement of the program-evaluation problem consists of a presentation of what the program evaluation expects to measure and compare.

Ideally, program evaluation should be concerned with significant concepts and those relationships of concepts which will improve programs. That is, the information generated from the evaluation should be cumulative to general knowledge and should aid in the evolution of more efficacious programs. It is not suggested that every program necessarily retain a university professor to thoroughly examine abstract concepts as they relate to program structure and function. However, program evaluations should examine practices or functions which represent universal concepts using accepted representative indices. Through this approach, one evaluation may be compared to another, and a more practical body of documented knowledge is made available to the field.

However, in some cases, one may see an index used such as, "the ratio of predicted Status 26 closures to achieved Status 26 closures." But what contribution has been made to the general body of knowledge concerning the efficacy of programs? The answer is none. Further, who says this predicted number represents an effective program? (See *Comparison of planned vs. actual performance*, pg. 22, for a discussion of this.)

To clarify the use of indices and the measurement of theoretical concepts, Chapter IV is a compilation of indices, and Chapter V contains examples of program evaluations using these indices to examine client outcome in relation to

broad theoretical concepts. (See *Structure and Dynamics of Social Intervention* pg. 95, *Rehabilitation and the Culturally Disadvantaged* pg. 96, and *A Conceptual Framework for the Description and Analysis of Human Service Delivery* pg. 101.) Other information which might be included in an introduction to the statement of the problem are findings in published data or references to similar program outcomes in similar jurisdictions.

Operationalization

In order to insure a viable methodological approach, the goals or aims of the evaluations must be carefully specified and defined in terms of the variables that will be examined and the operations that will be performed. In other words, the evaluator must specify in advance what he is going to look at, what he is going to measure and what comparisons he is going to make. This process of translating broad concepts into specific well-defined elements suitable for evaluation is called *operationalization*. The operationalized statement may be prepared in the form of a question or series of questions that an evaluator should be able to answer on the basis of analysis of program evaluation data. Several criteria that should be considered in the construction of these questions are considered below.

The question should be *precise*. Questions like "Why isn't our program as successful this year?" or "Have our clients changed for the better?" are so ambiguous and amorphous that they cannot be answered appropriately. Even the intent of these questions is unclear. A more precise restatement of the first question would be "Why have the number of Status 26 closures decreased over the last year?" Although this restatement is still not totally acceptable, at least we now know how "successful" is defined. The initial questions could have meant anything. Generally, then, the question must be stated so that there is no doubt as to its meaning or its intent.

Terms used in the question must be *well-defined*. Thus, in the question "Did we significantly increase the number of clients rehabilitated?" what is the definition of the term "rehabilitated"? Do we mean "closed status 26" or "employed at 6 months after closure" or "number who had a decrease in their welfare support?" What is meant by "significantly?" While some may consider this covered under "precision", there are many instances where a term is used which can be defined in the context of its use to eliminate possible alternative interpretations. In statements such as in the given example, each term should be defined as it is being used if there is any chance at all that alternative definitions exist. In many cases terms may be defined by referring to the actual measures (indices) used. Thus, the term "effectiveness" could be defined as the "Ratio of Acceptances to Closed, Status 26," the term "improved" could be defined as the average decrease in PA payments. As stated by McGuigan (61), "The main functions of good definitions are: (1) to clarify the phenomenon under investiga-

tion; and (2) to allow us to communicate with each other in an unambiguous manner."

The question must be *relevant* to the purpose of the evaluation. If an evaluation for effectiveness and efficiency only is being planned, then a question or index which relates to quality measures is not relevant. A question such as "How many status 26 closures did the project attain this year?" is not really relevant either, unless this index is compared to some other variable like, "number of 26 closures projected" or "number of clients accepted." (For cautions see *Comparisons of planned vs. actual performance* pg. 22). An index, which is a count of individuals that are part of a larger group, is not relevant for a definition unless the larger group is also included ("the count" = numerator, "larger group" = denominator; see Chapter IV).

The data required to answer the question must be *available*. Thus, if the definition of "rehabilitated" is the "number of clients in competitive employment after 5 years," the evaluator must have some assurance that either these follow-up data are already on file or can be obtained with reasonable effort. Such a question may be unanswerable because obtaining follow-up information five years after the client has been closed is extremely difficult, if not impossible, unless periodic contacts have been made in the interim. Further, if the evaluation requires filling out special forms or ratings etc., the cooperation of those required to do this work must be gained prior to implementation.

Finally, if the other criteria have been met, is the question worth the time, money and effort to find the answer? That is, will the answers be sufficiently important to warrant the *cost* involved in obtaining them. Here the evaluator must balance importance against cost. Some questions may be very important, but the cost involved in answering them is prohibitive. While other questions may not be as important, the cost for obtaining an answer is, at least, reasonable. The question "Has the VR program improved the quality of life in the United States?" may be a very important question (assuming "quality of life" can be defined) but the cost involved in obtaining an answer is probably so high that this question may never be answered. Other questions might be so trivial that they are not worth answering no matter how low the cost. The mental calculus of cost *versus* importance is a process that every evaluator must go through prior to setting up an evaluation.

Using the above criteria let us formulate one of the several questions that could be asked in the evaluation of a PA/VR Expansion Grant Project. Assuming that we are interested in the effectiveness of the program, the question could be stated: "Among clients on welfare, does the Expansion Grant Project which provides joint PA/VR case planning and joint provision of services produce a statistically significant higher ratio of "clients referred" to "clients closed Status 26" and a statistically significant greater reduction in welfare payments than does the regular Section 2 Program where no joint efforts are provided?" While this statement of the problem might seem cumbersome, it is reasonably precise. Terms have been defined on the basis of the indices to be used, the results of the

evaluation are generalizable because a concept is being tested, (joint service) and the variables between the two programs that are being compared are stated. Other questions concerning the program under evaluation could also be asked. Additional questions relating to the program's effectiveness, effort, efficiency, and quality would all add greater dimension to the overall evaluation.

Far from simply being a technical exercise, a well-drafted statement of the program evaluation problem represents an initial outline of the evaluation methodology. The operationalized questions are a major necessary step in the development of an evaluation plan. Development of measurable criteria from program objectives is a major use of operationalization.

SPECIFICATION OF METHODOLOGY

How will the program evaluator describe or compare the worthiness of the program under consideration? How will he develop a plan that is reliable, valid and free from bias?

Possibly the most important and difficult aspect of program evaluation is the selection and implementation of the evaluation methodology. The evaluator has at his command a sometimes staggering array of approaches, techniques, analytical methods and statistical tools. The critical question is, which of these will best solve the program evaluator's problem, contain costs, and consume a reasonable amount of time. A series of stepwise decisions is required, starting with the selection of the basic approach and ending with the selection of the appropriate mathematical tests. As stated before, rigorous application of sound methodological principles is the key element in the evaluation of any service program.

Below, we have attempted to present a general outline of the factors necessary for the development of an evaluation methodology. More comprehensive treatments of these factors may be found by consulting the references in the Reading List.

Approaches to Evaluation

As mentioned in Chapter I, there are two basic evaluation approaches: the research approach and the systems approach. Both have many variations with varying degrees of accuracy and cost. They represent two different ways of thinking about problems, goals and programs. Briefly, the research approach utilizes an inductive way of thinking. This means, specific facts, combined in some way, will show a relationship from which a general theory can be drawn. The systems approach is deductive in that it takes a general overview of a program, breaks it into specific units and then draws conclusions.

Research Approach

In many cases the research approach uses two different population groups to compare the effects of one type of service or program against another type of service or program, or one type of service against no service. The population group that receives no service or the non-experimental type of service is called a *control group*. Hatry (41) has suggested five evaluation designs that provide reasonable comparisons. These are:

Before vs. after program comparison, "Compares program results from the same jurisdiction measured at two points in time: immediately before the program was implemented and at some appropriate time after implementation . . . This design often is the only type which is practical when available time and personnel are limited. It is also appropriate when the program's duration is short, of narrow scope, and when the conditions measured have been fairly stable over time."

Time trend projection of pre-program data vs. actual post-program data. "Compares actual post-program data to estimated data projected from a number of time periods prior to the program." "This design is useful where there appears to be an underlying trend (upward or downward) over a period of time that would seem likely to have continued if the new program had not been introduced." Careful consideration must be given to any other changes that might be altering this trend or will alter this trend.

Comparisons with similar jurisdictions or population segments not served by this program. "Compares data from the jurisdiction where the program is operating with data from other jurisdiction where the program is not operating." "This design protects somewhat against attributing change to a specific program when in reality external factors that affect many local governments are responsible for bringing about the change."

Controlled Experimentation. "Compares pre-selected, similar groups, some of whom are served and some of whom are not (or are served in different ways). The critical [aspect] in this design is that the comparison groups are pre-assigned before program implementation so that groups are as similar as possible except for the program treatment." Traditionally, this design has been used when change in client behavior is anticipated. (For a thorough discussion of employing any variations on this design, see McGuigon in the Methodology section of the Reading List.)

Comparisons of planned vs. actual performance. "Compares actual, post-program data to targets set in prior years—either before program implementation or at any period since implementation." This design establishes goals or target numbers of clients to be rehabilitated, or costs per rehabilitation to be expended within a given period of time and then examines actual performance relative to these goals. (See the summary of North Carolina Program Evaluation Plan in Chapter V.)

The major difficulty with this last approach is the determination of goals. If the number of clients to be rehabilitated is set too low, then the program will surely reach the goal and be "successful" as far as "effectiveness" measures are concerned. (It might not be as successful if efficiency is examined.) If, on the other hand, the numbers are set too high, the program will be a failure or only the very "best" cases will be selected for service so that the number of Status 26's can be maximized. Some effort must be made to insure that the planned performance is a realistic estimate of what the program can expect to accomplish.

Minimally, how the projected goals were established should be explained in enough detail so that the reader understands the reasons and justifications for using a specific number for a program goal. (See index no. 25, Chapter IV.)

In Standard Setting or Expectancy Setting, formulas are sometimes derived from previous program performance so that present program activity can be quantified and compared. Thus, means may be computed for several variables for previous time periods and then weights are assigned to these means. A global picture of performance is then developed from these means and weights. (See Chapter III.)

Clearly, the major function of the above evaluation designs is to measure the changes in performance of the program clients by identifying those changes that can be reasonably attributed to the program's services. (See Effectiveness Measures, Chapter IV). In order to evaluate efficiency and quality, the same types of comparison designs would be used, but different measures would be employed. (See Efficiency Measures, Chapter IV). In choosing an evaluation design, three additional factors must be considered: timing (retrospective or prospective); the funds available, and the accuracy desired. "The first four designs discussed above are progressively more expensive, with the fourth usually considerably more expensive than the others. The first three and the fifth can often be accomplished with but a very few man-months of analytical effort, the amount depending heavily on the amount of special data collection required. The fourth design will likely take many calendar months and possibly years . . . The designs [other than the last], are also progressively more effective in providing, by far, the most reliable results. The designs presented above are not either/or choices. Some or all of the first three [and the fifth] are in fact, often used together . . ." (41)

Using comparative points of time, evaluation may concomitantly examine technological change and/or a program process.

Technological change or change in program philosophy, or changes in therapeutic technique may have an impact on agency operation or clinical services. Following the introduction of these changes, the obvious question becomes, "Did this technological change make the program more effective, more efficient,

or were higher quality services provided?" The purpose of such an evaluation is to determine, first, whether the new procedure is actually being followed and second, that changes in program effectiveness can be attributed to this new procedure (all other factors being held constant). This type of evaluation answers the question, "Is procedure X in fact being done and, in fact, is this change making a better program?" or "We have just started to use screen-in, screen-out procedures this year. Now we have a higher rate of Status 26's. Did the counselors really use the screen-in—screen-out procedures and is this contributing to the higher closure rate?" (60) This approach represents a combination of process and outcome evaluation.

Systems Approach

The systems approach has developed rapidly because of the advent of computers. Prior to this, the processing and analyzing of the large volume of data required in this approach was not realistic. There are several popular analytic techniques used today. These include operations research, cost-benefit analysis, systems analysis, and systems engineering.

Operations research is the application of scientific methods, techniques and tools to problems involving the operation of a system so as to provide those in control of the system with optimum solutions to problems. (50) In other words, operations research is the application of scientific method to the problems inherent in programs for the purpose of finding solutions to these problems. The problems, because they center on operational aspects of programs, will consistently involve costs. The methodology used in operations research follows the scientific method. One of the most popular applications of the operations research approach is cost/benefit.

Cost-Benefit/Cost-Effectiveness Analyses are becoming increasingly popular as evaluative tools in various fields. They started as analytical tools in the decision-making process of defense strategies and policies, and have now extended into human service areas such as rehabilitation program evaluation, health services effectiveness and the like.

"Both cost-benefit and cost-effectiveness analyses have their foundations in the operations research approach. Cost-benefit analysis can be applied wherever program inputs (costs) and outputs (benefits) are both measurable in money terms. The cost effectiveness approach is applicable whenever the program inputs are measurable in monetary terms but program outputs are not.

"Broadly defined, the cost benefit approach is an analytical study designed to assist a decision maker in identifying a preferred choice among possible alternatives. The cost effectiveness approach may be defined as an . . . 'attempt to minimize dollar cost, subject to some mission requirement (which may not be measureable in dollar terms) or conversely to maximize some physical measure of output subject to a budget constraint.'

"CB/CE, as approaches to decision making, have their own share of limitations and virtues, like all other analytical tools. Their limitations are those inherent in all analyses of choice. While CB/CE are readily applicable in evaluating projects which give rise to a stream of purely private benefits and private costs, they become increasingly tenuous for projects which generate significant externalities. The greater the magnitude and range of externalities generated by a project the more difficult it is to apply CB/CE analysis.

"While some sort of social discount rate may be used in the evaluation of such projects, it is the choice of the discount rate figure that is problematic. The choice of a social discount rate is arbitrary and by sufficiently lowering this discount rate, any investment can be made to look profitable."(82)

The above problem of defining *costs*, *benefits* and *externalities* (spill-over effects) must be clarified. In order to define what money represents *cost*, an illustration will be given. If General Motors produces a car, the cost of producing that car includes expenditures for central management staff, sales, promotion, depreciation of capital investments, and those unsold cars that are rejected and cannot be sold. Applying this concept to vocational rehabilitation, costs should include program costs (services, administration, purchased services and overhead) services free to V.R., cost of recycling clients, and client earnings forgone. (23) *Benefits*, on the other hand are more difficult to estimate. The problem involves the question of how to estimate the future income of rehabilitated clients. Questions surrounding salary increases and therefore, increased tax returns must be asked. Other questions surrounding the *externalities* generated in the rehabilitation process must also be considered. These include such things as the handicapped father who cannot work, but following rehabilitation, can stay at home and supervise the running of the household, thus allowing his wife to support the family. "In addition to this major limitation in the application of CB/CE to social programs, other issues involve the necessity of 1) specifying meaningful objectives or measurement goals; 2) predetermining adequate measures of performance; 3) developing viable policy and program alternatives against which to measure the cost benefit and/or cost effectiveness of any given policy or program." (82)

"Despite the limitations and difficulties inherent in the CB/CE approaches, both have the potential for contributing to more systematic and efficient decisions relative to social program planning and implementation. They are one type of tool to help the decision maker to understand relevant policy and program alternatives, estimates of risks, and estimates of possible pay-offs associated with varying courses of action." (82)

This brief presentation points out the most important strengths and weakness of cost/benefits and cost/effectiveness studies. (For further discussion of this technique, extensive references are provided in the Reading List. For practical examples, see summaries on Cost/Benefit Studies in Chapter V).

Systems Analysis describes a program (or system) in terms of the structure of the system and sub-systems, and the processes by which the function of that system and sub-system are accomplished. 1) *Structure* is the organizational breakdown of a program by levels and components. For instance, it may be broken down into levels of responsibility—Commissioner, middle management, and the counselors who render services; or it may mean the system divided by programs, such as physically disabled, mentally ill, PA/VR, etc. Also included in *Systems Analysis* is a description of the community's expectation of the program and the program's expectations of the community. The analysis of this relationship is usually examined through communication, financial support, committee attendance, etc. 2) *Function* means the activities performed by that structural unit to accomplish program objectives. For example, the Commissioner might be responsible for policy formulation, goal setting and the over-all program responsibility. Baumheier (7) developed a structural systems analysis based on responsibility and then identified five functions for each structural unit. These five functions were: problem identification, policy formulation and goal setting, planning, program operation, and evaluation. 3) *Process* describes how the structural unit operates to perform its function.

In order to define the structure, function, and process of a system, consistent use of certain analytic techniques has been made. These include space and layout analysis, analysis of records, forms and reports, methods and procedures analysis, organization studies, and analysis of communication channels. The specific tools employed usually include flow charting, model building, definition of the basic resources of staff, their level of knowledge and information, money available, and the identification of what data or information is needed to reach the program's objective or goal.

Systems Analysis may range from description, to analysis, to forecasting or prediction of the feasibility of alternatives. (See *Analysis of Data* for a brief description of mathematical tools). The major difference between systems analysis and operations research is the scale of the project and the time projection. Operations research provides more immediate solutions to problems.

"One factor to be kept in mind, is that comprehensive quantitative systems analysis to evaluate programs requires large scale commitment of financial resources, a high degree of research sophistication, and a total program commitment to evaluate goals. The reality is that the most rigorous type of systems evaluation is not feasible in many, or most, applied social program situations. This is not to say that evaluation methodology should not be defined and carried out as rigorously as possible, or that power is not lost when standards are less rigorous. However, it is to suggest some comprehensive systems approaches are the only feasible approaches in many situations and that with care they can provide meaningful information to serve as a valuable planning and evaluation tool." (28) In Chapter V, four articles and plans have been reviewed that loosely fall into the systems analysis concept. Some provide systems oriented conceptual guidelines while others also include systems oriented analytical and measurement tools.

Systems Engineering—Is a broad term usually associated with the design of new systems and the redesign of old systems. Historically, systems engineering implied the arrangement of physical components of some system, such as the arrangement of computers in an information system. Presently, the term has been broadened to include the design of any system. The purpose of systems engineering is to design all program components to interrelate with one another and for the program to interrelate with the total system. For instance, if a new PA/VR project was being initiated, the systems engineer would want to co-ordinate the new program with all existing programs. The method employed to design a plan would be flow charting of clients, information, and staff effort, and the establishment of an information-decision system. (47)

Reliability, Bias & Validity

Methodology selected or developed to carry out a program evaluation must attempt to insure that the results of the evaluation are reliable, free from bias, and valid.

Reliability is the agreement between two or more examinations of the same activity or process. This may be determined by having two or more judges observe the same thing and noting their degree of agreement (inter-judge reliability) or by using the same procedure on two different but equivalent occasions, or through the use of repeated procedures, or even using repeated evaluations and noting how well the results agree. Statisticians have more specific measures of reliability, but they all boil down to one thing—stability of a procedure to produce consistently similar results under constant conditions. If nothing external changes, any measuring instrument (be it a micrometer or a program evaluation) must yield consistent results. Reliability does not imply that the results are meaningful—only that they are consistent.

Bias is the tendency of a procedure or an individual to consistently favor a specific outcome of an evaluation. That is, does the evaluator or even the process ask and answer questions that will favor a specific point of view? Bias may occur through faulty methodology, by accident or by design. The problem of bias can pervade every aspect of the research from the statement of the evaluation objectives, the asking of research questions, the selection of the sample, the construction of the data collection forms, and ultimately, in the analysis and interpretation of the findings. It is only through the most rigorous application of sound methodology, by an evaluator who maintains an objective (let the chips fall where they may) attitude can this difficulty be obviated. In some cases, it is difficult for administrators to admit that their programs are ineffective or inefficient, but only through unbiased evaluation can positive program change be initiated.

Validity has been defined as the measurement of what one intends to measure. In other words, does the methodology selected for the evaluation

actually answer the questions posed by the evaluation problem, or, "Do you find out what you want to find out?" Those involved in statistics and test construction define several different types of validity in their work. For the purposes of program evaluation, the most meaningful of these is what is called "content validity." In practical terms this means examining the content of the evaluation program in order to determine how well this content will meet evaluation goals. As is stated in Isaac and Michael, "Thus, evaluating the content validity . . . for a particular purpose is the same as subjectively recognizing the adequacy of a definition." (48) A thoroughgoing examination of the indices and procedures involved in the evaluation is really the only way an evaluator has of insuring some reasonable level of validity.

Selection of the Sample

The population (universe) to be studied should be identified as clearly as possible, and if it is not feasible to reasonably study the entire population, efforts should be made to observe scientific sampling procedures. (84) In other words, what group of people, with what characteristics, should be studied to answer the questions posed? For example, you may wish to study a group of clients with specific characteristics, such as Vocational Rehabilitation clients with AFDC stipends who are under 35 years of age, or you may want to study the effectiveness of counselors with advanced degrees. Because of limitations of costs and manpower, the total group is often too large to allow for the proper collection and analysis of all data. Therefore, a small, representative group, or sample, must be selected.

Random Sample—individuals are selected from the population on a purely chance basis—That is, selection may be made by 1) pulling every fifth file folder for a 20% sample or 2) use may be made of a table of random numbers. If a sample is drawn that does not allow every member of the population an equal opportunity to be in that sample, then it is biased and not truly random.

Stratified Sample—A selection is made such that the sample adequately represents all pertinent variables in the specified population. For example, if it is determined on the basis of information from the Bureau of Census that the population of a given area consists of twenty percent black, half of whom have less than a high school education, the selection process then must result in a sample containing twenty percent black, half of whom have less than a high school education. This process must be repeated for all other pertinent variables in such a way that the sample selected provides a complete profile of the population.

The difference between random sampling and stratified sampling is that in the former it is expected that the characteristics of the population are reflected in the characteristics of the sample on the basis of probabilities, while in the stratified sample, the group members are specifically selected to represent the population.

COLLECTION OF DATA*

Data collection is often considered merely a dirty job in program evaluation (or in any type of analysis). However, it probably consumes the greatest amount of time and effort. The data to be collected is determined by the following: 1) Program objectives and evaluation criteria [indices] suggest what data can best indicate program effectiveness, 2) The evaluation design indicates specific time periods and for what specific groups data are required. Data used in an evaluation must have two characteristics: 1) They should be reasonably accurate and complete, 2) Since evaluation relies on comparisons of data collected at different times, and possibly for different population segments, data should be comparable. Many problems relating to the accuracy and comparability of data can be avoided if the evaluation is planned before program implementation. Pre-planned evaluations have another major advantage. Often data needed in a thorough evaluation are not normally collected. Certain data may not be obtainable in the "before" status after program implementation. With preplanning, provision can be made to collect such data both before and after implementation. For discussion purposes, we have identified five sources of data:

- Existing records and statistics
- Interviews with clients which in some cases involve citizens at large
- Rating by professionals
- Other special data collection procedures
- Experiences of other governments

Existing Records and Statistics

Some data are routinely collected on every government program. When these data relate to program objectives, they can be useful in evaluation. Often, statistics based on the data will be inadequate and it will be necessary to check basic records to generate the needed information. [See Sampling, Reliability, and Validity.] A special problem occurs in program evaluations that require tracing individuals beginning with their entry into the government system (such as reason for departure and subsequent success of the government-provided service). This need usually occurs in evaluations of treatment programs. Information on specific individuals from one agency must be "linked" with records on the same individual in other agencies. When done for a number of individuals, such information can be aggregated to relate disposition and treatment success

*Reprinted by permission from *Practical Program Evaluation for State and Local Governments*, Harry P. Hatry, Richard E. Winnie, and Donald M. Fisk. Copyright © 1973, The Urban Institute.

to reasons for entry and client characteristics—usually vital information for program evaluation.

The problem of confidentiality inevitably will be raised. Adequate steps are needed to assure that the information obtained on individuals will be safeguarded and used only to present aggregate summaries.

Evaluators must watch for changes in procedures or definitions of data categories during the collection of data used. This is especially important when comparing data from different jurisdictions. In such cases, if appropriate adjustments are not or cannot be made, the data are likely to be inconsistent.

Interviews with Clients and Other Citizens at Large

In many cases, interviews will be the only reasonable way to obtain real feedback on program effectiveness.

Three type of uses are described below:

- Status of individuals after leaving government sponsored programs.
- Citizen satisfaction with service.
- Certain “factual” data. [See Davis and Jacobs in Reading List and see section on Methodology.]

Interviews generally are conducted by written questionnaire, by telephone, or at-home (in person).

There are a number of limitations to the use of citizen surveys.

- There is considerable time and cost involved.
- Special technical resources and effort are needed to design the survey procedure and the questionnaire to be used.
- Survey information is limited by the respondent’s memory and willingness to answer.
- Unless a survey is conducted *before* program implementation—to be compared with results of a survey taken *after* implementation (using the same wording)—a direct before and after comparison will not be possible.
- Some citizens may refuse to be interviewed.

Ratings by Professionals

In some situations, subjective ratings by professionals, systematically obtained, may be appropriate for evaluating program effects. When used, the ratings should be output oriented and conducted systematically. [See Validity, Reliability & Bias]

Some form of pre-specified scale should be established with explicit

directions provided for using each reading on the scale. [See Sampling, Reliability, Validity & Bias.]

Finally, it is not appropriate that the individuals who give the service also provide the effectiveness ratings. Raters should be selected who do not have a self-interest in the outcome [see Bias].

To be meaningful within the context of evaluation, three factors should be standardized: the characteristics evaluated by professionals, the rating scale applied to these characteristics, and the conditions under which the ratings are made.

This method requires professionals who are competent to make judgements about the particular situations and who can be objective in their appraisal. Also, if a scale is not readily available, considerable time and effort using appropriate expertise will be needed to establish an acceptable rating system. The costs of making ratings could be large because of the time required for each observation and the specialized personnel involved. In some situations, as time passes, raters may tend to deviate from the rating scale. Periodic checks of ratings and retraining in the use of the scale when needed can alleviate this.

This method of data collection is basically subjective in nature and normally should be used in conjunction with more objective measurements.

Other Special Data Collection Procedures

The methods that might be used are limited only by the ingenuity of the evaluators. They might involve actual systematic field observations such as measuring the response time of police patrol, fire, or ambulance units.

Another approach might be called "special events analysis." This applies to programs that are particularly important in affecting unusual events on which information is not recorded.

A related approach has been labeled the "critical incident technique" which attempts to identify the number and type of critical incidents.

Special data collection procedures invariably entail added cost. Being "special," it may be impossible to recapture the "before" information unless the evaluation is planned prior to program implementation.

Experience of Other Governments

Information on the experiences of other governments will sometimes be useful for evaluations. On-site visits probably will be needed to screen out general public relations type of information and attempt to obtain comparable data. A second major problem is to carefully consider situations in the other government (such as significantly different clientele characteristics) that would affect the comparability of the jurisdictions.

ANALYSIS OF DATA

Levels of Statistical Measurement

The nature and organization of the data collected often dictates the level of statistical measurement (type of number scale)* and the consequent statistical tests that may be applied to describe and analyze the program data. The four levels of measurement are nominal, ordinal, interval, and ratio.

Nominal Scale—"Measurement at its weakest level exists when a number of other symbols are used simply to classify an object, person, or characteristics. The scaling operation is categorizing a given class into a set of *mutually exclusive* subclasses".(91) Thus, for example, all PA clients might be grouped into *mutually exclusive* classes of AFDC, APTD, OAA, etc. Nominal basically means "naming" and the scale "names" the group to which an individual belongs. It is important to remember that an individual may belong to only one class. All classes must be independent.

Ordinal Scale—ranking scale—the data may be ranked in order of magnitude along some dimension. The degree of the magnitude is not specified, and not necessarily uniform. The statement simply is, A is greater than B without knowing how much greater. Counselors may be ranked for performance: good, better, best; or closure 26's may be ranked indicating full-time employment is better than part-time employment, and part-time employment is better than sheltered employment. How much better is not specified.

Interval Scale—"When a scale has all the characteristics of an ordinal scale and when in addition, the distances between any two numbers of the scales are of known size, then the measurement is considerably stronger than an ordinal scale." (91) For example, in intelligence tests, an IQ change of 20 points at IQ 70 is approximately the same as an IQ change of 20 points at IQ 100. However, since this type of scale has no true zero point as its origin, it cannot be said that IQ 100 represents twice the intelligence of IQ 50.

Ratio Scale—when a scale has all characteristics of an interval scale plus it

*Number scales are mathematical entities and are not to be confused with behavioral scales, such as Scale of Rehabilitation Gain, Minnesota Measure of Employment Satisfaction, etc.

has a true zero point as its origin. Some examples are, dollars, number of clients, time. For example, an increment of \$5 when added to any dollar amount always represents an equivalent increase. In addition, we know that an expenditure of \$200 is twice an expenditure of \$100. (91)

Data Analysis Techniques

This section has been divided into two parts, mathematical techniques suitable for the research approach and mathematical techniques suitable for systems approach. In many cases, some of the same techniques can be used in both approaches.

Research Approach

One of the most firmly established analytical techniques for research revolves around the use of statistics and probability theory. Modern statistical techniques provide useful tools for the decision-making process. One of the major purposes of statistical analysis is "to provide methods of treating data so that the maximum information can be obtained with a predetermined risk of drawing false conclusions. The use of statistical methods is based on a reasonable assumption that accepted principles of logic and probability should produce correct answers more often than guessing." (50) However, statisticians and managers must be "constantly on the alert to check results against logic. Raw answers sometimes can be misleading, as in the case of spurious correlation. Results of statistical analysis which are to be used for estimation of prediction must be scrutinized for reasonableness." (50) The broad area covered by the term *statistics* encompasses a number of areas and topics. We shall consider the areas of descriptive statistics and statistical inference here.

Descriptive Statistics

Descriptive statistics concerns itself with collecting, tabulating, summarizing and presenting data for the purpose of describing a population or program. Its basic usefulness is the reducing of large masses of data into a meaningful form. There are three popular statistical techniques consistently used: frequency distributions, measures of central tendency and measures of dispersion.

Frequency Distribution is the first step in reducing data into a comprehensible form. For instance, the ages for 500 clients cannot be understood unless some sort of classification is done. Usually, this classification takes a tabular form:

Age	No. clients	% clients
1- 5	10	2.0
6-10	20	4.0
11-15	40	8.0
16-20	100	20.0
21-25	200	40.0
.	.	.
.	.	.
.	.	.
etc.	etc.	etc.

Once the frequency distribution is developed for each variable of interest (age, race, income, disability type, length of time on PA, etc.) the results may be presented in the form of *graphs*, *tables* and *density maps*. The presentation, when thoughtfully devised, may reveal a pattern, particularly when care has been taken in setting up categories for which two populations are compared. The major defect in using this approach as an analytical tool is that data can be severely distorted by accident or by design.

In most cases, further reduction of masses of data is desirable in order to describe them by their most important characteristics. Most often used are the measures of *central tendency* and measures of *dispersion*.

Central Tendency attempts to summarize a mass of data by characterizing the "typical" or "average" value of the data.

Mean—"the arithmetic average"—it consists of adding up all the values and dividing by the number of values added.

Median—when a set of numbers is arranged in order from highest to lowest, it is that number located in the middle such that one-half the numbers in the set fall below it and one-half the numbers fall above it.

Mode—that number in a set of numbers that appears most often. This measure is rarely used in most social research.

The following is an example of these three measures:

1
4
(5) - mode
(5)
7 - median
9
12
15
<u>20</u>
9/78=8.66 mean

Rate—a proportion of some variable—(see frequency distribution table, % clients)

$$\text{Rate} = \frac{\text{no. of members of population exhibiting a specific characteristic}}{\text{size of population}}$$

This may be also calculated as a rate per 100, per 1,000, per 100,000 etc.

Measures of Dispersion (variation) show how the data are “spread out” (distributed, dispersed) around the mean. That is, how closely the numbers in the set cluster around the mean. The most common measures of dispersion are the *standard deviation* and the *range*.

Range—commonly defined as the highest and lowest value in a set of numbers. This measure of dispersion is subject to many influences and is therefore, rarely used. It may have some value when the set contains few numbers. The range in the example above is 1-20.

Standard Deviation—is the most generally accepted measure for determining how the numbers in a set of data cluster around the mean. That is, the standard deviation of a set of numbers is small if the numbers are bunched closely around the mean and is large if the numbers are spread over considerable distances from the mean. This measure has the advantage of being the foundation for other more sophisticated measures in parametric statistical techniques. (See pg. 37).

Inferential Statistics

Inferential Statistics is based on probability theory and is an *inference* about a population when the data at hand covers only a small sample from the universe (population). Even if we observe an entire program's case load for all public assistance clients, that observation is regarded as a sample at that point of time. Therefore, these observations are sample measures, and sample measures produce estimates of characteristics of the entire population in the caseload.

There are two fundamental aspects of statistical inference—point estimation and hypothesis testing. The techniques involved are based upon results of mathematical investigations into the appropriate applications of probability theory to a given situation. (119)

Point Estimation—when one uses a sample measure (sample mean, sample standard deviation) as an estimate of the corresponding measure in the population from which the sample was drawn, the basic question is, “How good a guess are you making?” The answer must be based on knowledge of the manner in which sample measures may vary from sample to sample. This, in turn depends on the nature of the distribution characteristics. Statisticians have developed

suitable techniques for estimating population parameters. One of the results of these techniques takes the form of a *confidence interval* giving a specified confidence coefficient.

For example, a 95% (0.95) confidence interval gives two boundary values between which we can have a stated level of assurance that the population value lies.(119) That is, given a sample mean of 10 and a 95% confidence limits of 5 and 15, we may infer that 95 times out of 100 the population mean will fall somewhere between 5 and 15. Specific mathematical procedures for establishing the confidence interval of population include various measures of standard error, e.g., standard error of the mean, standard error of the rate, etc.

Tests of hypotheses are usually conducted when one wishes to know if a treatment or experimental condition applied to a group of clients has changed this group in some way. For example, we may wish to know if a certain counseling technique used with one group resulted in more rehabilitations than in another (matched) group in which the technique was not used, or we may wish to know if the provision of a new service has changed the behavior of a group of clients by comparing behavior before and after the service was introduced.

Prior to discussing the various tests of hypotheses, we must make a clear distinction between a research hypothesis and the null hypothesis. "The research hypothesis states the expectations of the researcher in positive terms. It identifies the variables or conditions which, in casual relationship, will be advanced to account for the results and is often devised from a theory".(48) Thus, if we use one of the examples cited above, the research hypothesis might be stated, "The introduction of group counseling techniques will result in a greater number of clients closed rehabilitated."

On the other hand, the null hypothesis is the one actually tested statistically. "It is an arbitrary convention hypothesizing that any relation or difference in findings [relation or differences between groups or treatments, etc.] is due to chance or sampling error."(48) Basically, this says that there is no true difference or no true relationship between groups or variables. Statistical tests basically test the null hypothesis. When the results of a test indicate the null hypothesis should be accepted, the outcome is that there is no statistical significance and any observed differences or relationships are probably due to sampling error or chance. When the null hypothesis is rejected, the results are statistically significant and observed differences or relationships, are probably "due to some determining factor or condition other than chance."(48)

The level of significance is basically that level of probability at which the results could be due to chance and at which the null hypothesis will be accepted or rejected. While no hard and fast rule exists for setting this level, one conventional level often used is .05 (5%). This translates as "Five times out of one hundred such a finding is due to chance variation." [For a more complete discussion see (48)].

An extremely important point must be made here. *Statistical significance in and of itself is not sufficient to demonstrate the practicality, usefulness, or effect of a program, service technique, or other intervention.* It is quite conceivable that a very small absolute difference between two groups can attain statistical significance. However, the question remains, does this difference have any practical significance. That is, while it is possible for an intervention that resulted in two additional status 26 closures to reach statistical significance, there is little practical meaning of this result in most cases. Note should be made that while it is possible to attain statistical significance without practical significance, it's rare for the results of a formal research study to attain practical significance without statistical significance.

Certain statistical tests require some assumptions to be made concerning the characteristics (parameters) of the distribution of the numbers in the sets of data to be compared. One such assumption is that the numbers in the groups to be compared are normally distributed. (Many readers will remember discussions of the normal or bell-shaped curve from their courses in elementary statistics). The type of statistics used to test hypotheses when such assumptions are made are called parametric statistics. There are times when one doubts if the assumptions concerning the characteristics of the distribution are indeed met. A second kind of statistical measure called non-parametric or distribution-free statistics may then be more appropriate. In addition, some experts have maintained that ordinal scale and nominal scale data lend themselves best to non-parametric statistical analyses while ratio scale data is best subjected to parametric analyses. Interval scale data fall into the grey area where both approaches are considered to be relevant.

Of course, the above is vastly over simplified. Many qualifications and constraints are inherent in the use of either approach and there is no clear agreement between experts on their use. (For an excellent discussion of the theory and specific non-parametric techniques, see books by Hays, and by Siegel in Methodology section of Reading List).

The following list of measures is limited to the most popularly used techniques in inferential statistics. Each measure is followed by a brief description of its use. It is not the purpose of this section to serve as a statistical text. The use of any given measure in a specific application requires a certain degree of statistical sophistication on the part of the user to determine applicability. Misapplication of a measure can sometimes lead to more difficulties than not using a statistical measure at all!

t-ratio or t-test is used to measure if any differences found between the means of two variables were due to an intervention, service, technique, etc., or were due to chance or sampling error.

Analysis of Variance is used to measure whether any differences found

between two or more variables were due to chance variations. Analysis of variance can become very complex, handling large numbers of groups and subgroups simultaneously. The results obtained from this measure break down any observed differences into variables and combinations of variables that give rise to these differences. This is a powerful and often used measure.

Chi square is a non-parametric measure used to determine interdependence or association between data grouped in categories. For example, looking at status 26 closures, one may wish to determine if the race of the counselor and the race of the client had any effect on closures, assuming clients were assigned randomly to counselors:

		RACE OF CLIENT	
		White	Non white
RACE OF COUNSELOR	White	No. closed Status 26	No. closed Status 26
	Non-White	No. closed Status 26	No. closed Status 26

Simple correlation measures the degree of association between two sets of numbers. Thus, by use of this technique, one may wish to measure the degree of association between years of work history prior to rehabilitation and the salary received after successful closure. *Correlation coefficients* vary between +1.00 and -1.00. A correlation of +1.00 indicates a perfect positive association, e.g., for each year of prior work history the starting salary after closure increased by \$500. A correlation of -1.00 indicates a perfect negative association, e.g. for each year of prior work history the starting salary after closure decreases by \$500. A correlation of 0.00 indicates no association between the two variables. Intermediate values of the correlation indicate greater or lesser degrees of positive or negative association. It is very important to note that degree of association does not necessarily imply causality. In other words, one cannot assume that the years of prior work history caused the increase (or decrease) in starting salary after closure, rather one can only conclude that they vary in unison in some manner.

One of the important uses of any correlation technique is that of prediction. If a significant correlation (any correlation may be tested to determine if it is significantly different from a zero correlation) is obtained, one set of numbers may be used to predict a portion of the income after closure from merely knowing how many years of prior work history the client had.

Partial correlation measures the relationship between two variables when more than two variables are present by mathematically holding one or more variables constant. Thus, in the example above, it might be discovered that those clients with a greater number of years of work history are also older than those with lesser work histories. The examination of the relationship between work history and salary is said to be contaminated by the age variable. (In a like manner, any examination of the association between age and income would be contaminated by the work history variable.) Partial correlation allows one to measure the association between work history and income with the variable of age held constant (usually stated, "age partialled out").

Multiple correlation measures the relationship between one set of numbers and the optimally weighted combination of two or more other sets of numbers. Generally, multiple correlation is considered to be the correlation between a criterion (dependent) variable and the combination of two or more predictor (independent) variables. For example, this measure can be used to indicate the correlation between the combination of educational level and years of prior work history (predictors) and salary after status 26 closure (criterion).

Regression coefficient is the measure used in the actual prediction of one set of numbers from another. It is closely related to the correlation coefficient, but differs in that it is directly used for prediction.

Multiple regression equations are closely related to the multiple correlation, but are used to measure the actual contributions of independent variables toward the prediction of a criterion. For example, using this technique, one might find when trying to predict Final Salary After Closure that Years of Prior Work History contributes to a certain amount, X; Educational Level adds another amount, Y; Number of Months in Active Status adds another amount, Z; etc. and the combination of all the independent variables predicts the criterion at level R.

Factor analysis is a measure for analyzing patterns of intercorrelations between many variables and attempts to identify elements which account for the patterns. In examining successful closures, one may have a substantial number of variables. Correlations are then computed between each variable and every other variable. This results in a table of intercorrelations (intercorrelation matrix). After factor analyzing this matrix, one might discover that there are certain elements that account for the patterns of correlations, e.g., one element might be Work History which might include the variables of; Number of Years Worked, Level of Skill Required in Past

Work, Salary Earned, Level of Responsibility, etc. Another element might be Educational Level which might include the variables of Number of Years of Schooling, Type of Courses Taken, Scores on Standardized Achievement Tests, etc. Factor analysis is a difficult measure to use and interpret, but, when correctly applied, can supply valuable insights into the dimensions or elements (factors) which make up a given construct.

Systems Approach

The mathematical techniques used with the systems approach include the techniques mentioned in the research approach plus some additional complex techniques more suited for the analysis of large scale systems. This approach, in particular Operations Research, uses symbolic models as "mathematical descriptions of an activity which expresses the relationship among various elements with sufficient accuracy, so that it can be used to predict the actual outcome under any expected set of circumstances." (50)

There are three categories or circumstances in which models are developed: 1) the solving of large scale complex problems under certain conditions, 2) the solving of complex problems under uncertain conditions, and 3) the simulation of systems for experimentation and evaluation in predicting the consequences of change in policy, conditions, etc., or to verify or demonstrate a new idea. Specific mathematical techniques used in each of the above three categories include:

- Large scale complex problems under *certain* conditions—linear programming, input-output analysis, quadratic and dynamic programming.
- Large scale complex problems under *uncertain* conditions, queuing theory, and Monte Carlo Technique.
- Simulation—all of the above mathematical techniques plus scheduling simulations, Critical Path Analysis and PERT (Program Evaluation & Review Technique).

In the following paragraphs brief conceptual descriptions of the mathematical techniques are presented. For a comprehensive discussion and presentation of formulae, see the section in the Reading List on the systems approach (Ackoff, R.L., Fundamentals of Operation Research).

Linear Programming—This mathematical technique is applied to the so-called "mixing" or allocation of resources problems. That is, following the specifications of a goal which is not yet attained, such as rehabilitation of 900 clients, the planning for the allocation of resources (staff, funds, facilities) to different operations, (duties, expenditures and space) in such a way as to either minimize the total cost or maximize the total return.

For example, "How can more clients be treated with a given budget or how can the budget be allocated between staff and operations to produce more status 26's?" This mathematical technique aids in the selection of the most desirable course of action from a number of available courses of action. "The [administrator] must consider a large number of factors affecting his decision(s). Also, these factors are interdependent so that the [program director] must consider them individually and in relation to each other. Finally, the choice must be made of one solution or course of action from among many obvious alternatives and, perhaps, some others which are not so obvious. Without linear programming techniques, these decisions must be based on experience, feel, intuition, and hope." (50)

The limitation of this technique is the absolute necessity for a definition of clear-cut objectives. That means overall program objectives. "Different departments often pursue divergent purposes because each follows its own objectives, whether reduced costs, higher production, (or) greater utilization of capacity. Within the linear programming framework, such diversity will not work. The analytical procedure makes it painfully clear that to push for one goal is to do so at the expense of another. From the viewpoint of the larger system, only one of these purposes can be the correct one for the system under the set of conditions which exist. Thus, linear programming provides a rationale for the homogeneous operation of those systems to which it can be, and is, applied." (50)

The second limitation is that there must be a straight line or linear relationship between important variables. Linear programming cannot deal effectively with more than one set of conditions at a time. Optimal solutions to real problems involving simultaneous change in several variables are extremely difficult to compute. (1,37,50)

Input-Output Analysis was developed as part of a linear program. It provides a quantitative framework for the description of an entire economy or system. Basic to input-output analysis is a unique set of input-output ratios for each production process. For example, when a given number of clients are needed to produce a given number of rehabilitations, it is possible to predict exactly the necessary flow ratios at referral, evaluation and training in order to reach the program objective. Given the construction of such a matrix, estimates could be made of the impact of a decrease in counselor working time due to inservice training, or lowered working hours. This technique is expensive to employ and would work most effectively on a national level. (1,50)

Quadratic and Dynamic Programming—"These two techniques have been developed in order to offset some of the disadvantages inherent in linear programming."(50) This is an algebraic technique, but unlike linear programming, it can handle nonlinear relationships. The objective of *dynamic* programming, as with linear programming, is that of facilitating solutions to sequential problems. It is a method of solving step by step case flow problems. "It is a method of solving multistage problems in which the decisions at one stage become the conditions governing the succeeding stages."(50)

For instance, quadratic or dynamic programming would be used for smoothing of production levels to meet season job demands, or the allocation of limited resources between current output (900 clients) and say, retraining of staff to more productive counseling techniques. This technique is somewhat developmental and expensive. Again, it would best be applied on a national level.(1,50)

Queuing theory is a well developed analytic technique which seeks to apply probability theory to determine the length of waiting lines (queues) for a service. It seeks to balance service available with services required. The technique provides a methodology for computing the average length of a waiting line as a function of services. Queuing theory may be applied in determining the least costly number of services required. In short, the application of queuing theory permits the quantitative balancing of cost of waiting time of clients against cost of additional services. (1,37,50)

Monte Carlo Technique—There are times when a number of the variables that go into the solution of a problem involve some uncertainty. Or some of the events that may happen can be stated only in terms of their probability of occurrence. In more complex cases, the problem involves the interaction of a number of events, each of which has some differing probability of occurring. In cases where these events cannot be represented by standard probability curves, the Monte Carlo technique can often be applied to obtain some solution to the problem.

Basically, the Monte Carlo technique develops a data base for problem solving through the use of random sampling. A random number generator is programmed to produce numbers according to the probabilities involved. These numbers are then used for problem solving as if they were actual observations of the real world. The numbers are, in a sense, approximations of actual data and as such, can be used to find approximate solutions to problems that otherwise could not be solved by the more rigorous mathematical approaches.(37,50)

Critical Path and P.E.R.T.—The fundamental principle observed in planning any major project whether it is processing clients or building a house, is that about 15 percent of the jobs in the project actually determine the ultimate time required for the entire project. It then follows that any speeding up or delaying of these critical jobs will affect the duration of the process. Critical Path and Program Evaluation and Review Technique (P.E.R.T.) are planning and analysis techniques which identify these key activities. From this, a flow diagram is constructed to represent the process with the key activities identified. The critical path establishes the longest sequence of activities in a network to establish the minimum length of time for accomplishment of the end event of the project. The path from the first event through the end event consists entirely of critical activities and is therefore called the critical path. Limitations of PERT techniques include the inability to specify which time elements of the process are not predictable. Secondly, it is not practical for activities which are in a re-occurring cycle, that is a flow diagram usually only covers one cycle and therefore only gives the average time for processing that cycle (average processing time of a PA client). (1,50,112)

INTERPRETATION OF THE DATA AND CONCLUSIONS

There are no hard and fast rules for interpreting the results of data analyses, and there are certainly no rules for drawing conclusions from these results. If the evaluator followed a well planned and objective procedure, certain findings may lend themselves to obvious, unitary interpretations. However, other findings may fall into a "grey" area where alternative (and sometimes divergent) interpretations are possible. It is at this point that the experience of the evaluator becomes paramount. In many cases, he must rely on his personal knowledge of the specific program under examination and its components to interpret his data and come to some conclusions. In other cases, he may confer with program personnel or "significant others" in order to clarify his findings. In any case, he must always be able to justify and defend his interpretations and conclusions in light of his objective data. In addition, it is incumbent upon the evaluator to resist having his findings distorted in order to "document" preconceived notions of program or agency personnel.

Program evaluation is only as good as the accountability built into it, according to this author, who has been in programs under evaluation as well as having played the role of evaluator. The accountability program evaluation he describes here requires three components: goal setting, feedback, and real consequences.

III

THE NINTH PANACEA: PROGRAM EVALUATION

by
Robert A. Walker

Ever since man created organizations, he has energetically sought to make them more effective in meeting client needs. Program watchers have viewed the application of a string of so-called panaceas, beginning with exhortation and moving on to multi-disciplinary teams, indigenous staff, consumer participation, planning units, coordination schemes, research utilization, and the perennial political favorite, federal dollars. While each application has helped to some degree, those that receive human social services will probably maintain that

delivery systems have a good long way to go. Supposedly, the service consumer should take heart, because poised on the horizon comes THE solution to decaying delivery systems: program evaluation.

Before we prepare to bask in a sudden responsiveness of organizations to client needs, it might be well to recognize that program evaluation has been with us for a long time, beginning with "evaluation research," which may date back to the Civil War era. Despite those efforts by a small band of rational thinkers and the hundreds of programs that have been diligently evaluated, program mediocrity has stubbornly persisted. In effect, pass/fail, or summative, program evaluation by itself has not quite succeeded in creating more effective organizations.

A more recent version of program evaluation arose from the views of federal policy-makers, who proclaimed that process or input measures were "out" and outcome measures were "in." As a result, a new version of program evaluation was developed and labeled formative. This model defined the purpose of program evaluation as to improve the organization; it called for the development of feedback. No longer were results to be viewed by policy-makers and University biostatisticians alone. The data were to be carefully, if not helpfully, displayed to program managers and staff to "improve program performance." The result? Perhaps some momentary discomfort for program staff, but annual budget requests were still fulfilled and for most organizations it was business about as usual.

Notwithstanding the experiences of others, the author pursued a parallel course while employed in a voluntary employment agency. During that time, three demonstration projects were developed and fitted with a series of program evaluation measures. The technology was well done and produced a nice series of reports, which chauvinistically suggested that although the benefits of the projects were rather minimal, the experience was undoubtedly worthwhile to all. Apart from the traditional platitudes, two insights were gained from this effort. First, program evaluation did not improve the benefits over the life of the projects and secondly, the two programs that showed positive results were discontinued while the third, an abysmal failure by the evaluation measures used, was expanded.

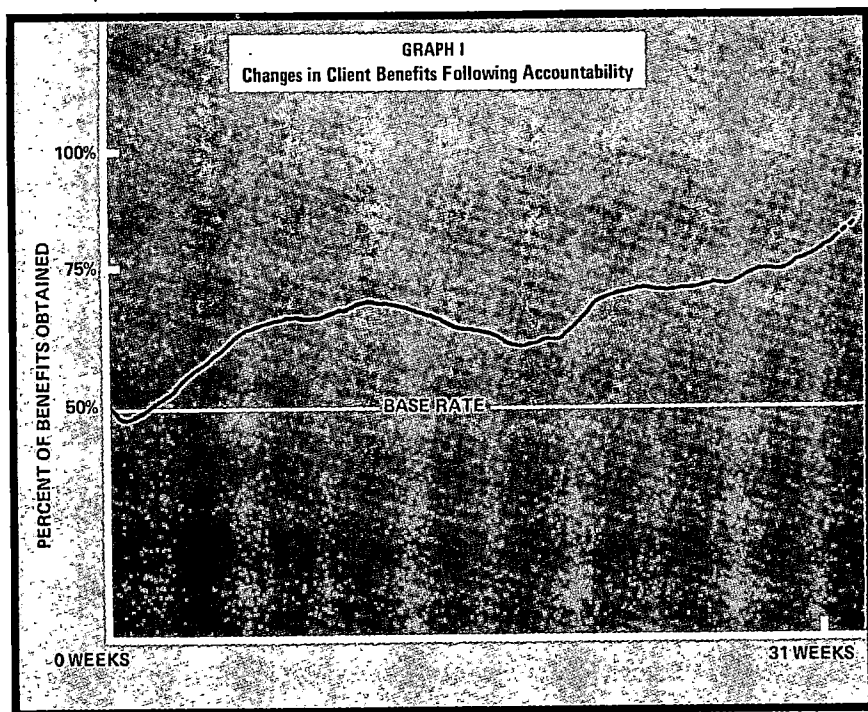
An analysis of just why program evaluation failed to work a much-needed magic resulted in two conclusions. First, from a technical standpoint, the measures and feedback provided incompetent data for the program manager and staff who delivered services. Second, and perhaps most important, there were no consequences provided within the organization that would result in the development and maintenance of desirable staff behaviors. The typical programmatic response could be described as, "Damn the data, full speed ahead."

Using the stimulus of accountability rhetoric proposed by John Gardner and the technical application of learning theory, this author in quiet desperation developed and installed a third version of program evaluation, labeled as

accountability program evaluation, in a voluntary employment agency. Graph I shows the results.

The actual figures revealed that the overall benefit increased 31.6 percentage points over a 31-week period. A doctoral thesis completed by Orville Kiel, using this data, concluded that the gains were statistically significant and that the experimental variable, accountability program evaluation, was successful in achieving program goals.

Although the methodology used was in many ways similar to traditional versions of program evaluation, the major difference in accountability program evaluation can be summarized in this way: "Accountability program evaluation requires that evaluation data be of such quality that each staff member's contribution to achievement of the program goal is clearly known, so that consequences can be equitably and differentially provided." The remainder of this article will describe how the author is developing various forms of accountability in several diverse systems in hopes that the material will stimulate readers to experiment with his "third version of program evaluation."



As is the case with other forms of program evaluation, the business of accountability program evaluation begins with goal-setting. Enough already has been written about the problem of vague goals, non-goals, and inappropriate goals. In addition to those nightmares, accountability advocates must face the issues that not only is a program measure needed, but that each person who contributes to that goal must also be measured and that levels of expectancies must be clearly set. The complexities involved in designing such a system are described in the following section.

The author was recently involved in an effort to develop an accountability system within a local program of a national manpower program designed to deal with the hard-core unemployed. Using a base rate assumption that most programs don't know where they are heading, the first task was to analyze the data reports currently being used to measure program effectiveness. The existing data report was designed to provide feedback to staff, which enabled them "to determine the overall effectiveness of the program by providing summary statistics on program activities." It did so by summarizing outcomes and transactions for all enrollees terminating the program during each report month. On one single sheet of paper, 246 scores were displayed. In addition, the local program staff developed a supplementary performance report and added to the data base an additional 51 measures. In effect, each program month, project management and staff were deluged with 297 program indicators in hopes of "determining overall effectiveness of the program. . . ."

Following recovery from this flood of stimuli, it was apparent to staff that the data did not tell them which of the 297 measures of "total performance" were the most important. Staff recognized that the smorgasbord of offered choices created a virtual playground for the capricious value judgments of the program managers. As all indicators were apparently of equal value, it would be possible for fickle management to select a new indicator each month that need not be repeated until 24 years had passed. The abundance of goal choices created so much staff anxiety that monthly review sessions resembled a group version of Russian roulette. The problem of determining program performance became even more confusing when the question was asked: "What is acceptable performance for each of the 297 measures?" If different levels of expectancies were set for all of the 297 measures by each of three administrators, the game of roulette could be played until retirement—a possibility that became more attractive to staff as each program month passed. It was apparent that if accountability were to be introduced, a clear definition of program success was an absolute necessity.

Accountability program evaluation began by doing away with process indicators (e.g., 90 per cent of all enrollees should receive "Orientation") because, apart from their questionable validity, they tended to concretize the delivery system. If feedback were to self-correct the system and help accomplish goals, then the delivery system needed to be structured in such a way that changes

could take place. In order to define the major program goal, a set of 21 alternative sets of outcomes were presented. These, along with "implications for a given choice," were offered to both management and supervisory staff. This resulted in the selection of three major goals that could ultimately be transformed into measures. However, since several goals were to be used, it became necessary to weight these individual goals so that they could be combined into a composite score representative of overall program "success."

The need for a single composite score is easily illustrated by the problem created in having multiple measures. Table I shows simulated data on the three measures used to measure effectiveness by the three teams. Try answering the question, "Which team deserves a letter of commendation?"

TABLE I			
	Team A	Team B	Team C
Employed 90 Days	23%	16%	32%
Hourly Wage	\$2.50	\$3.10	\$2.25
Program Costs	\$2,612	\$1,852	\$3,462

If the question is difficult to answer with only three measures and three teams, consider the plight of the program manager faced with 297 measures and a staff of more than 60. The solution to the problem of which is the best team requires that weights be attached to each of the three measures. Then, and only then, can rewards be equitably provided. The specific weight or value for each of the three major program measures was obtained by asking this question: "Assuming that 100 points represents all possible values, how many points would you attach to each of the three program goals?" In a surprisingly rapid fashion, the value of each measure was attached as follows: Continuous Employment of 90 days (50 value points), Hourly Wage (10 value points), Program Costs (40 value points). As part of the development of an overall system, management's decisions were translated into a table of program values that not only represented an explicit definition of management's value judgments, but were also to be used by each staff member filling out a form at the time a program participant terminated the program. In effect, feedback began when each staff member filled out the form. A representation of this form is shown in Table II.

Each of the forms completed for individual enrollees at the point of termination was routed to a unit that aggregated them in various combinations for management to view. However, before the reports could be optimally used, it was necessary to define aggregate levels of expectancies so that staff were aware not only of what types of goals were to be achieved, but the expected rate, or

TABLE II
Table of Program Values: Benefit Calculation

Enrollee's Name: _____ Termination Date: _____
 Circle level of achievement _____

POINT SCHEDULE	EFFECTIVENESS:		EFFICIENCY: PROGRAM COSTS	POINT SCHEDULE
	EMPLOYMENT STATUS	HOURLY WAGE		
50	Employed 90 days			50
40			Less than \$1,000	40
30	Self-employment		\$1,000 to \$1,999	30
20			\$1,999 to \$2,999	20
10	Employed part-time	Over \$3.00/hr.	\$2,999 to \$3,999	10
5		\$2.99 to \$2.00		5
0	Unemployed	Less than \$2.00	Over \$4,000	0

Points Earned: _____ Employment Status _____ Hourly Wage _____ Program Costs _____ Total Points: _____

Staff Signature: _____ Date: _____

level, of achievement. In this effort, three general levels were established: 1) Below minimum, 2) Optimal, 3) Achievable maximum.

Specific levels for each goal were set by staff judgment and an analysis of a sample of terminated clients, accomplished by retrospectively applying the measures and point values. The scores were then aggregated and linked to each staff member responsible for that benefit. This resulted in a distribution of staff performance. The expectancies were set so that the level of "optimal" was currently being achieved by approximately 20 per cent of the staff. "Minimum" was set so that approximately 15 per cent of staff fell below that level. The remainder of the staff fell in an intermediate range.

Use of a retrospective study not only provides a rational and defensible basis for rate-setting, but also enables a comparison to be made of the effectiveness of the approach because base rate data is available. Through use of the weighted measures, composite score, and the levels of expectancies, the data unit after minor statistical manipulations now produces for the manager a monthly report similar to the one shown in Table III.

The clarity of the goals, weights, and expectancies now makes it possible for the manager to provide a variety of consequences. Based on the monthly report shown in Table III, Team C should be rewarded for their overall performance and for high scores in two areas: employed 90 days and low program costs. Team A should also be recognized for their high performance on the employment goal. If the manager wishes to improve both overall performance or the performance of either Team A or B, he knows on the basis of the report the goal area where the performance problem exists. Team B must improve the number of people employed for 90 days, and Team A can improve their performance by reducing program costs. One method of intervening would suggest that management bring together in a staff training program those teams with high performance in a given area (Teams A and C, for employed 90 days) with staff from those units that score low in that goal area. In effect, both management and staff know for the first time who are the high or low performers, and it is possible for management to intervene with change agents to aid marginal teams.

Although a definition of explicit goals, weights, and expectancies, followed by the provision of consequences, can be expected to improve overall performance, the introduction of accountability should show that performance differences between staff decline as overall program performance improves. The reason for this requirement is simple. If two out of three teams are superb performers and nudge the total program into the "better than expected" range, the improvement is of no value to those clients who have the misfortune to be served by the marginal team. Graph II shows the changes in performance obtained in the voluntary employment agency (also charted in Graph I) over a 31-week time period.

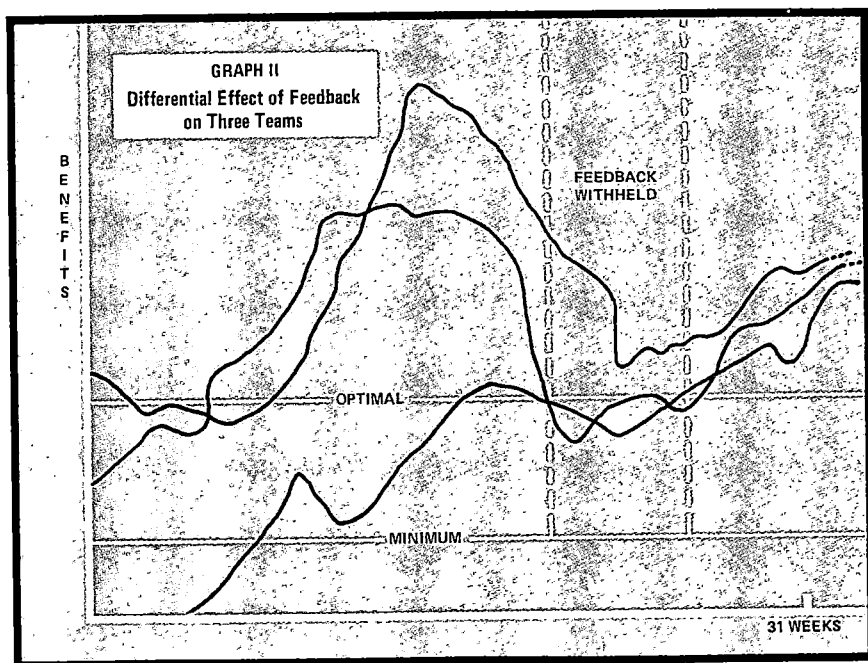
TABLE III
Monthly Performance Chart

Month: _____

Elements	Expectancies		Team	Team	Team	All Teams
	Minimum	Optimal	A	B	C	
Composite Score	40	60	52	43	73	56
Employed 90 Days	22	32	36	18	42	32
Average Wage	4	6	4	6	5	5
Program Costs	14	22	12	19	26	19

□ = optimal or above

○ = below minimum



The space between the vertical dotted lines reflects a time period when feedback to the teams was experimentally withdrawn because of a "Breakdown" of the on-line computer. Feedback was experimentally a powerful factor since performance declined during this period and increased once feedback was re-instituted. In effect, the graph shows that accountability program evaluation is performing at an optimal level as all teams became virtually identical in terms of results, with the composite of all teams exceeding the optimal.

Clear goals, weights (if goals are multiple), and various levels of expectations can stimulate staff in the general direction of program objectives. Eventually, however, a point of "technical over-stretch" will be reached, and regardless of rewards for good performance, an answer must ultimately be provided to the question: "Just how do I get better results?" Because recognition and social praise have their limitations, the task is to develop feedback that tells staff how desirable results can be improved.

A description of how to construct truly useful feedback reports cannot be dealt with fully in this article. Instead, some general comments are offered concerning the purposes of feedback, information users, general specifications, and examples of reports.

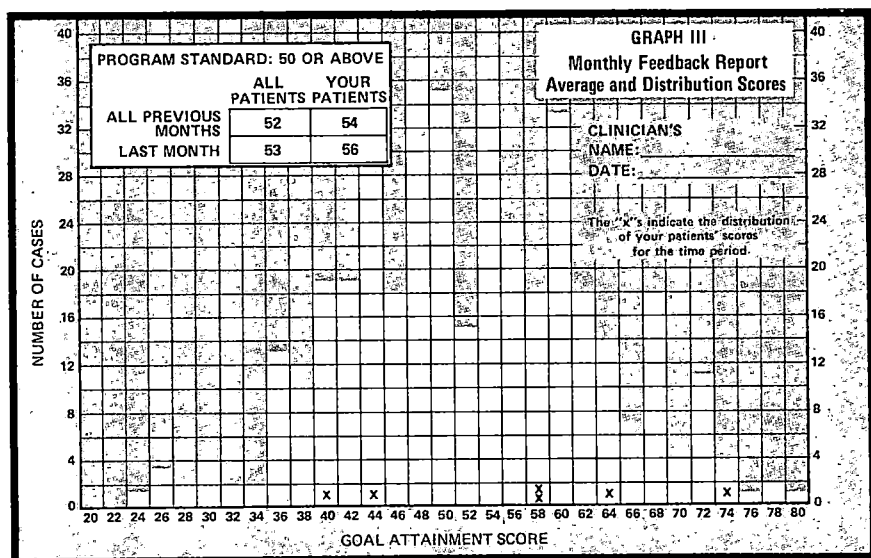
As is the case with most human endeavors, the first task of feedback is to determine its purpose. This typically involves more than simply improving outcomes and lowering costs. A common objective, for example, is to insure that

the organization is serving those for whom the program was intended. If this is an objective, feedback data must be supplied that monitors the degree to which conformance is maintained with that defined target population. Another common purpose is to monitor key portions of the delivery system, such as the volume of clients, program utilization data, and such constant headaches as insuring that written reports on clients are completed on time.

Once the objectives of feedback have been determined, the next task is to define all potential users of performance information. The three major groups of feedback recipients usually are: change agents, who make programmatic decisions and usually consist of managers, supervisors, or staff from support units, such as staff training and research; direct providers of service or line staff, such as caseworkers, counselors, or occupational therapists; and an often neglected group of "interested others" who support the system by allocating money, formulating policy, or by paying taxes. A fourth, but completely neglected user, is the service consumer. Although consumers are able to obtain performance information on cars, toasters, and restaurants, they seldom have access to data that tells them if an organization can help them achieve personal goals. Although this author has never been asked to supply such a report, creeping Naderism may some day require that such data be displayed so that consumers can select those organizations that offer the best result potential.

Specifications that describe feedback reports in terms of such dimensions as content, style, and frequency will vary with the type of agency and the characteristics of the staff. Regardless of these variations and the need to customize the reporting of results, some general observations might be made. First, most staff are rather impatient with reports and will seldom labor long and hard to understand them. The task of the feedback designer is to develop a report so that the reader knows and understands the data within a matter of minutes. This can usually be accomplished by providing a limited set of data in as clear a fashion as possible. Displaying information on relative performance against several standards is also useful. That is, feedback offers the opportunity to make several comparisons, which might include: a worker compared to an absolute standard, a worker compared with others, and finally, a worker compared to himself. This data is often organized along a time dimension so that "trend" information can be presented. Graph III shows how feedback information was presented to clinicians in a mental health center.

In this instance, the purpose of feedback was to improve the degree to which program goals were achieved. The program goal was a Goal Attainment Score, which represented how well patients achieved individually set personal goals. The display allowed the clinician to compete against a program standard (score of 50 or above), himself over time, or all other clinicians. One of the major problems involved in this effort was that the number of patient scores to be fed back each month was often quite low. In order not to mislead the clinician, data was supplied in a distribution format that told the clinician how many



scores were aggregated during a reporting period so he could discount the values based on the size of the sample. In any case, the report format was such that the clinician was presented with a variety of comparisons, one or more of which could stimulate an increase in performance.

The most difficult task from a technical standpoint is to provide feedback information to users concerning how program results can be improved. Feedback reports need to do more than show performance. If it is to be helpful, feedback needs to supply data that provides a clue as to what can be done to improve results. One approach is to use exception reports. That is, the concern is not with service recipients who "make it," but rather, feedback concentrates on those who are defined as exceptions—who "miss." Exception reports are difficult and time-consuming to develop and can be quite expensive to obtain. Yet, if technical plateaus are to be surpassed, they are probably an essential. An answer to the question as to why an exception occurs can be developed using at least three different types of feedback reports.

The simplest exception report informs the user as to where an exception happened. In certain types of agencies, especially those that have services structured in modules or components, the place or time an exception occurs can provide management with some useful data. Table IV illustrates an exception report that defines where a dropout or a "termination prior to a program benefit" occurred in a manpower program.

Since lessening the dropout rate can contribute greatly to achieving the overall program measure, this type of report at least can identify those components or places where a problem exists. It does not provide management with a definitive answer concerning the why of a failure, but as a minimum it narrows

TABLE IV
Exception Report:
Point of Termination

Dates: June 1, 1972 to Sept. 1, 1972		Number of Enrollees: 86
Component		Per cent of Terminations
Orientation		10
On-job training		12
Adult basic education		8
Job referral		26
Holds/"stop for"		16
Follow-up		28

the search down to a specific location within the delivery system.

A second type of report, the **who** of an exception, involves client characteristics. This report is constructed on the assumption that staff have clinical skills and if they are provided with clue data they can develop more effective ways of dealing with "misses." This feedback strategy calls for the development of a list of client characteristics that staff feel are related to the program goals and organizes those characteristics in a report. In the example of feedback developed for a mental health center, staff was first presented with a list of reliable data elements and from that list 10 items were selected that were felt to be most valuable. Each monthly feedback report presented data on followed-up patients for each of the 10 characteristics or activities. A copy of the report is shown in Table V.

By viewing the feedback report, the clinician is able not only to obtain a picture of low scorers, but also to compare "winners" and "losers." A by-product of this type of feedback report is that management can obtain a profile of those clients who benefit and those who do not and either develop better techniques through special programs or consider the option that such patients should not be admitted for services.

The third, and by far the most difficult, feedback report is the one that will answer the question concerning **why** an exception occurred. When properly done, such a report can provide highly specific data to program personnel, enabling them to correct a deficiency. A typical first thought in constructing such a report is that it can be developed by asking the staff why an individual, group of individuals, or type of failure occurred. Unfortunately, this approach usually results in one discipline responding with a vague term such as "poorly motivated" or "low ego strength," while behavioral types may be overly specific and state that the problem was "failed to seek work 4 times a day." The result is

TABLE V
Patient Analysis Report:
Goal Attainment Scores and Patient Characteristics

Therapist Name: _____

Date: _____

GOAL ATTAINMENT SCORES BELOW 50						ITEMS	GOAL ATTAINMENT SCORES ABOVE 50					
ACTUAL SCORES				40	44		58	58	64	74		ACTUAL SCORES
				3	5	Number of Therapy Visits	13	12	7	9		
				31	52	Age	18	26	41	24		
				M	F	Sex	M	F	F	F		
				no	yes	Medications (Pre)	yes	no	no	yes		
				yes	yes	Medications (Post)	no	no	yes	no		
				295.3	302.0	Diagnostic Code	296.2	301.8	301.8	302.0		
				D	D	Marital Status	M	M	M	SEP		
				yes	yes	Prior Therapy at MHC	no	no	yes	no		
				S/HA	A/BA	Shiple-Hartford	VS/VS	VS/VS	A/S	A/S		
				7	7	Pay Status	8	8	7	8		

that the decision-maker is faced with a long list of esoteric reasons that tell more about how staff think rather than how to develop intervention strategies. Perhaps a better approach is to develop a list of high frequency "whys" and to ask the staff which of them best describes an exception. If such a list has within it some behavioral observations or operational descriptions and an open-ended feature allowing for rare observations, then at least a partial answer to the "why" of a failure is possible. A representation of such an exception report is shown in Table VI.

As the reader will probably infer from this section on the "why" of feedback reports, the technology is definitely a non-science, and at best a very weak art. There is no question, though, that functional feedback reports need to be developed in such a way that they lead to program improvements. Unless feedback reports can reveal the why of failures, it may never be possible to fully optimize the benefits of human service systems.

While some will still maintain that it is love that makes the world go 'round, more recent thinking suggests that it is really consequences. Whether the term is accountability, incentives, contingencies, reinforcers, or rewards, consequences can play a powerful role in shaping organizational behavior. Although most administrators find goal-setting and feedback loops to be grossly unpleasant, they often look forward to providing consequences. In a field supposedly populated by those with a great concern for others, there are administrators with

TABLE VI
Exception Report

ENROLLEE'S NAME: _____ Date: _____

PRIMARY BEHAVIORAL OBSERVATION: Applied for work but never offered a job

COMPONENT WHILE TERMINATED: Job referral

POSSIBLE REASON (check all that apply):

- ☐ 1. Had marginal job skills
- ☐ 2. Could not explain problems (age, education, prison record, etc.) in an interview
- ☐ 3. Did not tell employer he wanted to work
- ☐ 4. Poor explanation of skills
- ☐ 5. Appearance a problem
- ☐ 6. Poor on application blanks
- ☐ 7. Other _____

Suggested Services: _____

Signed: _____ Date: _____

a punitive life-style who can hardly wait for feedback to identify staff laggards. They often see that the balance of power has finally shifted and management has recaptured program control. Better yet, if the benefit measures were developed by professional staff, the laggards can be hoisted by their own petards. However attractive such a possibility, managers need to be told that punishment is a poor technique. Although pleasant to administer for that fleeting moment, it will ultimately destroy accountability because avoidance of pain is a poor incentive. The optimal rule is as follows: reward good behavior and provide the bad with skillful neglect and bountiful supplies of technical assistance.

It might be pointed out that a suitable set of rewards for good performance can be provided in public programs, regardless of the more than considerable constraints imposed by civil service. Staff are also people and such dispensations as social praise and recognition can and should be used. They can be provided through informal ways ("Good job") or through such formal mechanisms as staff meetings, newsletters, letters of commendation, and so forth. Other attractive possibilities are to allow time off, attendance at conferences, "self-improvement" days, and merit salary increases. In effect, there are more incentives available in "non-incentive" systems than most people would anticipate. They do need to be clearly defined and provided under proper sets of circumstances, best based on some general principles of learning theory.

The introduction of contingencies into the land of no consequences does create not only some staff discomfort but a technical problem for those installing such systems. The technical problem is that when accountability exists, and obtaining results becomes important for staff and program survival, ways can be found that will improve benefits without any of them actually accruing to the client.

As most managers of a human service know, there are many ways to get benefits, if they are needed. Perhaps the most popular is to be highly selective concerning who receives services. By careful avoidance of recidivist-prone clients, such as drug addicts, alcoholics, or public offenders, many programs appear to be scoring high on program benefits. The task of the program evaluator is to require management to specify target populations in such a way that benefit gains are not obtained through changes in client selection. A second alternative sometimes employed by result-oriented organizations is to reduce the volume of clients served, thereby saturating the service recipient with professional staff. This can be controlled by defining volume requirements (i.e., the number of clients to be served in relation to the capacity of the system) or, better yet, by holding the program accountable for costs. A third method of obtaining spurious benefits involves the point in time when benefit measures are applied. For example, one program solved its performance problem by not formally enrolling clients until they had passed the high dropout period. With dramatic suddenness the program became an overnight success. A similar strategy involves measuring outcomes before the client loses them. This is often used in some manpower

programs that measure the stability of employment after a sparse 30-days. An opposite approach is to measure benefits as late as possible, in hopes that the environment will achieve the goal. A final method used to obtain spurious benefits is to manipulate goal-setting. For those organizations that use gain scores based on pre- and post-measures, setting goals that are easy to achieve or defining client status-at-entry as lower than it actually is can spuriously increase program benefits. The need to plug such loopholes following the introduction of consequences is not to suggest that staff cannot be trusted. However, it is the responsibility of program evaluators to make sure that when benefits scores do increase, the benefits actually do accrue to the clients.

It may come as no surprise that many people are dissatisfied with the performances of human service systems. The time has clearly passed when service recipients will be humbly grateful for the leavings passed out by condescending organizations. Nor will imagery work, as consumers are not at all impressed by professionalism, computers, and great shows of concern. People seek services from organizations to achieve personal goals and organizations too often fail to help. Although the solution of accountability program evaluation will surely not be the last nor necessarily the best of solutions, it can create organizations that self-correct themselves in order to achieve explicit goals. While the method is a technical problem, the real issue involves the willingness of organizations to accept accountability as a way to improve performance. Although Francis Bacon once said that time is the greatest innovator, it is doubtful that human service organizations as they now perform have much time left. As uncomfortable as are explicit goals, feedback, and consequences, accountability coupled with program evaluation can provide an opportunity for service consumers to achieve a reasonable quality of life. Most would agree that they need it, deserve it, and one day shall get it.

IV COMPENDIUM OF MEASURES

The compendium of measures is a categorized list of representative indices culled from over 75 studies and articles concerning evaluation of Vocational Rehabilitation Programs. This organization of the categories was developed from the program evaluation matrix described in Chapter I, pg. 10. That is, indices were divided into three program levels—general program, service delivery and program management, and then sub-divided into the categories of effort, effectiveness, efficiency, and quality. This list of the major categorical breaks was then reviewed by a number of consultants knowledgeable in each aspect of program evaluation in Vocational Rehabilitation. The review process was addressed to the general topics of reliability, validity and power of each index, plus the consultants' general opinion of the index. Through this process, it was hoped that the sum of the combined comments would help to establish the utility of each index for field use.

Note should be taken that an index may be used to represent one measure in one part of the compendium and another measure in a different part of the compendium. For example, index 2a, Program Cost or VR Budget is listed as a program-effort measure and yet index no. 125, Budget for Fiscal Admin./Total VR Budget is an efficiency measure. The explanation is obvious, the same index may be used to measure different things when combined with another index or depending on the goal of the program level.

Indices may be constructed in a variety of ways. These include simple counts, simple ratios, relationships of ratios to fixed numbers or averages, and the multiplication of ratios. Some of these are illustrated below in Table I. While

some ratios may be computed as percentages, others do not lend themselves to this reduction and should be presented as a decimal fraction. Whether to present a ratio as a fraction or as a percentage is largely a matter of logic.

A reasonable convention when constructing ratios is to set up the denominator to represent a total group and the numerator to represent that portion of the total group which is of interest. Thus, the ratio of Status 26 Closures to All Closures would be expressed:

$$\frac{\text{Total No. Status 26}}{\text{Total No. Status 08+26+28+30}}$$

where "Closed Status 26" is a part of the group "All Closures," or

$$\frac{\text{Administrative Budget}}{\text{Total VR Budget}}$$

where the Administrative Budget is part of the Total VP Budget. Such ratios can vary between zero and 1.00 (0-100%). Where the ratio "cost per unit" is to be computed, costs should appear in the numerator.

Table I

Index	Description	Equation and Example	Result
A	Fixed Index (Number, Count, Average, etc.)	-----	100 closed Status 26. Average Difficulty Index (ADI) = 4.05
$\frac{A}{B}$	Simple Ratio	$\frac{\text{Total VR Budget}}{\text{Annual No. Status 26}} =$ $\frac{\$10,000}{100} = \$100$	\$100 per Rehab.
$\frac{A}{B} : C$	Simple Ratio relative to Fixed Index	$\frac{\text{Total VR Budget}}{\text{Annual No. Status 26}} : \text{ADI} =$ $\frac{\$10,000}{100} : 4.05 = \$100 : 4.05$	\$100 per Rehab. with ADI of 4.05
$\frac{A}{B} \times \frac{C}{D}$	Two simple ratios multiplied	Acceptance Ratio x Rehab Ratio = S.S. Rate $\frac{\text{Annual No. Status 10}}{\text{Annual No. Status 08+10}} \times$ $\frac{\text{Annual Status 26}}{\text{Annual Status 26+28+30}} =$ $\frac{400 \times 100}{600 \times 400} = \frac{40,000}{240,000} = .16666$	Service Success Rate = 16.67%

COMPENDIUM OF MEASURES

INDEX

PROGRAM

EFFORT

COMMENTS ON THE VALUE OF THE INDEX

Indices of program effort are inputs, or the staff, budget, clients and facilities used by a program to do its job. In and of themselves, these indices are meaningless unless compared to standards, or similar programs in similar jurisdictions, or used in establishing indices of program efficiency.

$$\left[\text{e.g., } \frac{\text{inputs}}{\text{outputs}} = \frac{\text{effort}}{\text{effectiveness}} = \text{efficiency} \right]$$

1. Annual No. of Clients
 - a. annual no. status 08
 - b. annual no. status 26
 - c. annual no. status 28
 - d. annual no. status 30
2. Annual Costs
 - a. program costs (administration, purchased service, overhead, salaries)
 - b. services free to V.R.
 - c. cost of recycled clients
 - d. earning forgone by clients
3. Total No. of all staff

- 1.
2. These are useful indices that must be considered in estimating cost of rehabilitating clients.
- 3.

COMPENDIUM OF MEASURES

INDEX

EFFECTIVENESS

Client Behavior Change

4. Change in welfare status
(14,39,92,99,105,118)
5. Average Grant Amount Decrease
(14,39,64,114)
6. Had Job at Closure (Status 26)
(31,34,56,92,105,110,113,114)

COMMENTS ON THE VALUE OF THE INDEX

Indices or program effectiveness are outcome and follow-up findings. They indicate how well the program met the goals of changing current client behavior, reaching expected levels of program performance (e.g., 5000 Status 26's per year), reaching target population, change in test response, etc.

4. This is a useful index. There is a question of its sensitivity as a measure of reduction of economic dependency, since it is an "all or none" measure. It does not allow for meaningful comparisons between jurisdictions because of differing welfare regulations.
5. This index is a more sensitive measure of the reduction of economic dependency than 4. Caution must be observed when using this index because of ongoing welfare payment changes due to legislative mandate. Such a change could contaminate study results.
6. This is universally one of the most popular indices for measuring program effectiveness. However, improvements could be made if skill level were included (occupational status, DOT skill level). Also note must be made that it fails to consider improvements made in the non-vocational functions (externalities).

$$7. \frac{\text{Annual No. Status 26}}{\text{Annual No. Status 28 + 26}} =$$

(56)

$$8. \frac{\text{Annual No. Status 26}}{\text{Annual No. Status 26+28+30}} =$$

(56)

$$\left[\begin{array}{l} \frac{\text{rehabilitated}}{\text{served}} \\ \frac{\text{rehabilitated}}{\text{accepted}} \end{array} \right] = \text{rehab rate}$$

$$9. \frac{\text{Annual No. Status 26}}{\text{Annual No. Status 08+26+28+30}} =$$

(56)

$$\left[\begin{array}{l} \frac{\text{rehabilitated}}{\text{all closures}} \\ \frac{\text{rehabilitated}}{\text{referred}} \end{array} \right] = \text{rehab rate}$$

7-9. Since the ultimate Vocational Rehabilitation goal is employment, these are useful measures of total program effectiveness. In order to insure validity of these indices, all clients closed 08, 28, and 30 must be followed-up to determine reasons for failure. A disproportionately low rate may indicate ineffective service or inappropriate referrals.

COMPENDIUM OF MEASURES

INDEX

10. Annual Status 20's Transferred
to Status 26 : Employment Rate
of Community
11.
$$\frac{\text{Annual No. Status 10}}{\text{Annual No. Status 08 + 10}} \times \frac{\text{Annual No. Status 26}}{\text{Annual No Status 26 + 28 + 30}}$$
12. Annual Income Increase
(39,44,92,114,118)
13. Monthly Income Increase
(39,93,110,113)
14. Annual Household Income
(39,92,110)

COMMENTS ON THE VALUE OF THE INDEX

10. Experimental. This index measures program effectiveness as it relates to economic conditions in the community. Gives no indication of the quality of client's rehabilitation.
11. Service Success Ratio = Acceptance Ratio X Rehabilitation Ratio (Michigan VR)
12. & 13.
This is a useful economic indicator when work-related costs and inflation are taken into consideration. A defect lies in the inherent assumption of *zero* capacity for earning if unemployed at referral.
14. This index may be helpful in determining the value of a program when the rehabilitant frees other family members for work. If index is used for that purpose, note must be made of other environmental factors also.

15. Hours worked per week
(39,64,92,105,110)

16. Job Retention
(44,105)

17. Job Benefits
(44)

18. Type of Job Held
(45,64,92,105)

19. Change in Skill
(64,92)

***Response Change to Standard
Tests***

20. Heimler Scale
(114)

21. Life, Stress and Psychological
Functioning
(77)

15.

16. Universally, this is considered one of the most important indices. It indicates stability of rehabilitation and probable job satisfaction. Some caution must be noted however, as retention of the same job may reflect lack of career ladder opportunities.

17. If data are available, this index could measure the quality of the job itself.

18. There is no definition of standards for job quality; skills vary with individual jobs and thus are unrelated to title.

19. Difficult to assign values to different skills. May be helpful when changes are from lower skill to higher skill (e.g. machine operator to machine group foreman).

Note: Changes in client *responses* to tests do not necessarily reflect changes in client *behavior*.

20. Experimental.

21.

COMPENDIUM OF MEASURES

INDEX

- 22. Job Satisfaction and Family Functioning
(77)
- 23. Rehabilitation Service Satisfaction Questionnaire
(70,100)
- 24. Scale of Rehabilitation Gain
(111)

Expected Levels of Program Performance

- 25. Annual No. Status 26
Projected No. of Status 26 Per Year
(73,76,113)
- 26. Establishment of Viable Program
(44)

COMMENTS ON THE VALUE OF THE INDEX

- 22. Not widely used in Rehabilitation. It is often misused in evaluating program effectiveness, because lack of job satisfaction leads to job termination. Family function difficult to measure.
- 23. Gives clients' view of services effectiveness. Experimental.
- 24. This scale attempts to assess effectiveness by assimilating diverse factors. Experimental.
- 25. This index does not take into account the "weight of the closure" (i.e. case difficulty). The index only says that "if we produce a certain number of rehabilitations per year, we are effective". Unless substantial documentation is presented to justify "the projected No. of Status 26's" this is arbitrary and not a true measure of effectiveness. The goal may have been set too low to challenge the program or set so high it is unattainable.
- 26. Although this index has been used, the word "viable" must be operationalized to be quantified.

27. No. of Jobs Developed + Promoted
by Program
(46)

27. Although this index represents the goal of many programs, it does not mean that jobs increased in a community, but rather the number of jobs made available to the agency increased. This sort of unspecified goal promotes a "revolving door" of jobs & workers. [Displacement (q.v.)]

Target Population Reached

28. % of Target Population Reached
(44,70)

28. & 29.

This is a useful index, however, it is difficult to reliably estimate the size of the target population.

29. $\frac{\text{Annual Status 26}}{\text{Size of Target Population}}$
(44,70)

30. $\frac{\text{Newly Disabled Population}}{\text{Total Population}} \cdot \frac{\text{Status 10}}{\text{Status } 08 + 10}$

30. This is a useful index indicating the ability of the program to keep up with new need. [incidence (q.v.)]

31. Annual Status 26 Per 10,000 pop.
(109)

31. This is a useful index if the prevalence of disability of the program's population base is known. [prevalence (q.v.)]

32. % of Institutionalized Clients
De-Institutionalized
(8)

32. This index does not represent a Vocational Rehabilitation measure of effectiveness unless related to vocational placement.

COMPENDIUM OF MEASURES

INDEX

PROGRAM

EFFICIENCY

33. $\frac{\text{Annual VR Budget}}{\text{Annual Status 26's}}$
34. $\frac{\text{Annual Cost}}{\text{Annual Status 26}}$
35. $\frac{\text{Annual VR Budget}}{\text{Annual No. of Status 26}}$
: Average Weighted Closure
Index No.
36. $\frac{\text{Annual No. Status 26}}{\text{No. Professional Staff}}$

COMMENTS ON THE VALUE OF THE INDEX

Indices of Program Efficiency are $\frac{\text{effort index}}{\text{effectiveness}}$ It indicates how economically a program met its goal.

33. This is a useful index of efficiency. Note should be made that the budget only includes appropriated funds for the program—(services, purchased service, administration, overhead)
34. Cost per Rehabilitation. Similar to Index 33. However, it includes Annual VR Budget plus service free to VR, such as Medicare, Welfare, Board of Education. A deficit of the index is the problem of clients carried over from one fiscal year to next. While this is a mismatching of some funds, it generally offers an average.
35. This is a proposed index based on the use of Weighted Case Closure. This index compares cost of an average rehabilitation to case difficulty (see Weighted Closure pg. 85)
36. This can also be compared to the Average Weighted Closure No., as above.

QUALITY

37. Attitude of Client to Agency
(66)

38. Rehab Services Satisfaction
Survey
(46)

39. Measure of Employment Satisfaction
(103)

40. Measure of Employment Satisfactoriness
(104)

41. Job Relatedness to Vocational Training
(116)

42. Structured Case Review
(72)

43. Pre and Post Treatment Structured
Interview
(100)

Indices of Program Quality indicate conformance to accepted standards of professional practice and acceptability of service to the client.

37. This index is difficult to measure. Negative attitude may reflect slow or poor service. However, negative attitudes may also reflect client militancy.

38. This scale measures one aspect of program quality—client's opinion on "how good" he thought the program was. Since the counselor has more direct contact with the client, the results tend to measure how well the client liked his counselor.

39. An accepted established scale measuring client satisfaction with his employment.

40. An accepted established scale measuring the employer's satisfaction with the client.

41. A scale measuring relevancy of training to employment.

42. This method establishes certain standards as criteria for "good" counselor performance and then examines one counselor's case records, or a program's case records for meeting the standards.

43. Potency and value entirely dependent on how well the interview is planned and conducted. (experimental)

COMPENDIUM OF MEASURES

INDEX

44. Adherence to Regulations and Professional Standards

SERVICE DELIVERY

EFFORT

45. Total No. of Services
46. Total No. of Purchased Services
47. Annual Cost Per Service
48. Total No. of Targeted Projects by Disability
49. Annual Cost of Each Disability Project
50. Total No. of Staff by Title

COMMENTS ON THE VALUE OF THE INDEX

44. This is best measured through use of a check list.

Service Delivery indices represent measures of the performance at technical level of the Vocational Rehabilitation Program. Specific components of service delivery are Vocational Counseling, Evaluation, Training and Placement. No attempt has been made to include indices for medical or psychological services. Service Delivery Effort indices represent inputs affecting all subunits.

45. Over emphasis on this index can lead to the development of inappropriate services.
- 46.
- 47.
48. This is useful only in conjunction with demographic information (e.g., no. targeted projects to target population).
- 49.
- 50.

EFFECTIVENESS

The purpose of the indices of Service Delivery Effectiveness is to measure change in client behavior.

These have been covered under PROGRAM EFFECTIVENESS INDICES: Change in Client Behavior : Indices 4 through 18.

EFFICIENCY

51. Total Service Cost
Annual Status 26's

51. This index is a measure of the efficiency of the service component of the program. *It cannot be used as a measure of total program efficiency.*

52. Annual Status 26's : Average
No. Professional Staff : Weighted
Closure
Index No.

52. This index is useful only when compared to standards or to other similar programs and jurisdictions demonstrating similar case difficulty.

QUALITY

SEE PROGRAM QUALITY INDICES—Indices 37 through 44.

Counselor Performance

The Vocational Counselor is a component of the service delivery level. Coordination of the services to the client are his responsibility and his effectiveness and efficiency is dependent upon effectiveness and efficiency of the evaluator, the job trainer and the placement officer's work. Therefore, indices of counselor performance must only concern activities for which the counselor alone is responsible.

COMPENDIUM OF MEASURES

INDEX

COMMENTS ON THE VALUE OF THE INDEX

EFFORT

- | | |
|---|---|
| <p>53. Time allocated to client contact, clerical, recording, reporting, etc. for each counselor by disability. (35,39,109)</p> | <p>53. There is conflicting opinion as to the usefulness of this index, with some viewing it as useless and others viewing it as useful. Those who feel it is useful think it indicates counselor's ability to get his job done.</p> |
| <p>54. No. of Client Intake Interviews</p> | <p>54. There are conflicting opinions concerning the usefulness of the indices related to number of interviews and ultimate effectiveness and efficiency of the counselor. Many feel there are so many related variables that these indices are useless. Monitors future case flow.</p> |
| <p>55. No. of New Clients Seen Per Month</p> | <p>55.</p> |
| <p>56. No. of Interviews Per Client (33)</p> | <p>56. Does not indicate depth of interview.</p> |
| <p>57. No. of Contact Hours Per Client</p> | <p>57. Some studies have found this to be the significant determinant for successful closure.</p> |

EFFECTIVENESS

58. $\frac{(\text{Counselor CDF}) \times (\text{No. Status 26's})}{400 \text{ or } (\text{Agency CDF})}$

58. Composit Difficulty Factor (C.D.F. = total of time in active status, average case service cost of 26 closures, average No. of case services per closure, and average Sermon Difficulty Index score). Experimental. May be computed on individual counselor basis or district or agency basis. Possible value for inter-district or inter-state comparisons. (34)

EFFICIENCY

59. $\frac{\text{No. Contact Hours Per Counselor}}{\text{Annual No. Status 26's per Counselor}}$

59. Experimental index of counselor efficiency—assuming random assignment of clients to counselors.

60. Rehabilitation Counselor
Rating Scale (71)

60. Experimental Index.

See PROGRAM *Quality* for scales that may be used to establish individual counselor performance quality.

Evaluator
Job Trainer
Placement Officer

To date, few indices have been established for evaluating the effort, effectiveness, efficiency, & quality of vocational rehabilitation supporting personnel. Those we have found are listed below.

EFFORT

61. No. of Industrial Placement Contacts

61.

COMPENDIUM OF MEASURES

INDEX

COMMENTS ON THE VALUE OF THE INDEX

EFFECTIVENESS

62. Placements Within Industrial
Contacts

62. While this is a useful index for measuring placement's effectiveness, note should be made concerning case difficulty and type of client training and type of available jobs. (Proposed index)

63. No. Placements

63. Outcome index sometimes used for placement. However, this index really reflects total program activities and as such should not be used as a measure of placement effectiveness.

PROGRAM MANAGEMENT

Specific indices measuring program management have been divided into the components of administrative functions, community relations, staffing, inservice training, and fiscal administration.

EFFORT

64. \$ Spent on Program Management

64.

EFFECTIVENESS

65. Profile of Organizational
Characteristics
(Authority Structure)
(53)

65. An accepted scale measuring effectiveness of the leadership process, motivation, communication process, goal setting, the control process, performance goals, and training.

66. Marvin's Management Matrix
(62)

67. Minnesota Satisfaction Questionnaire
(103)

66. An experimental scale measuring the effectiveness of the process of working through others to produce worthwhile results and the generation of usable ideas.

67. An accepted scale rating satisfaction of supervisors and counselors on 21 items measuring ability utilized, compensation, policy, practices, etc.

EFFICIENCY

68. Administrative budget
Total Budget

68. Proposed popular index used in industry to measure administrative efficiency.

Administrative Functions

Much of a program's effectiveness and efficiency depends on factors unrelated to the technical level, but rather to the working conditions of the staff which enable staff to perform their various jobs easily. Many attempts have been made to measure these "conditions". Indices do not always properly reflect the satisfactoriness of the working conditions provided by management and consequently many times rely on the professional judgment of the evaluator. Below, we have taken items from Texas' Handbook on Program Evaluation. Some of these areas are:

69. Allocation and Utilization of Office Space.

70. Procurement of Satisfactory Supplies and Equipment.

71. Availability of Job Descriptions. (76)

72. The Experience, Professional Memberships, and Licensure of Staff in Relation to Needs of Particular Jobs. (21,45)

COMPENDIUM OF MEASURES

INDEX

COMMENTS ON THE VALUE OF THE INDEX

- 73. Equitable Pay Scales.
- 74. Reliable and Useable Record Keeping (Case Records, and Administrative Records).
- 75. An Effective Communication System. (76)
- 76. Proximity to Transportation for Staff and Clients.
- 77. Elimination of Architectural Barriers for Staff and Clients.
- 78. Delegation of Decision Making (45,64).
- 79. Staff Turnover—Monitoring Index for Staff Satisfaction.

Community Relations

The goal of the component, Community Relations, is to procure inputs and assure mechanisms for disposition of the outputs. The majority of the indices in this subsection have been used at one time or another. However, many of the indices of effort, effectiveness and efficiency are newly classified and are therefore, considered experimental.

EFFORT

- 80. No. of Staff Committed to
Community Relations.

80.

81. \$ or Staff Time Used in
Community Relations.

81.

82. No. of Presentations Made to
Community Organizations.

82.

83. No. of Different Brochures Printed
per Year.

83.

84. No. of Press Releases Distributed
per Year.

84.

EFFECTIVENESS

There are two types of interagency agreements, those that refer clients to Vocational Rehabilitation (IN) or those that accept referral of Vocational Rehabilitation clients (OUT).

85. No. of Interagency Agreements

85.

86. Annual No. of Staffing Agreements

86.

87. Annual No. of Clients Referred to
VR via Interagency Agreements.

87.

88. Annual No. of Clients Referred
Out via Interagency Contracts.

88.

89. Annual No. of Acceptances by V.R.
Annual No. of Referrals to V.R.

89. Proposed index for effectiveness of referring clients to VR per
inter-agency agreements (appropriation of referrals).

COMPENDIUM OF MEASURES

INDEX

- 90. Annual No. Of Acceptances Out
Annual No. Of Referrals Out
- 91. Waiting Time by V.R. Clients for
Acceptance to Receiving Agency.
- 92. Annual No. of Printed News
Releases.
- 93. Annual No. of Staff Outstationed
or Placed on Other Agency Pay Roll.
(64)

EFFICIENCY

- 94. Annual Acceptance by V.R.
No. Agreements
- 95. Annual Acceptances by other Agencies
No. OUT Agreements
- 96. No. Printed News Releases
Annual No. New Releases

COMMENTS ON THE VALUE OF THE INDEX

- 90. Proposed index for effectiveness of referrals OUT via interagency agreements.
- 91. Proposed index for effectiveness of referrals OUT via interagency agreements.
- 92. Proposed index for effectiveness of public relations and community relations in general.
- 93. Proposed index for effectiveness of interagency staffing agreements.
- 94. Proposed index for efficiency of referring interagency agreements.
- 95. Proposed index for efficiency of V.R. third party agreements.
- 96. Proposed index for efficiency of PR department.

97. No. Staff Outstationed
No. Agreements

QUALITY

98. Referring Agreement (IN)
What Agencies and for What Purpose.
(76)

99. Referring Agreement (OUT)
What Agencies and for What Purpose.
(76)

100. Inter-agency Staffing Agreements.
What Agency and for What Purpose.

101. News Releases to Whom and
What About

102. News Releases to What Media and
for What Purpose

Staffing

EFFORT

103. Annual \$ Spent in Recruiting.

97. Proposed index for efficiency of interagency staffing.

98.-100.

Proposed indicators for quality of the interagency agreement.
These indices are judgmental in nature.

The goal of this component is to procure qualified staff in as short a time as possible.

103. & 104. Proposed Index.

COMPENDIUM OF MEASURES

INDEX

COMMENTS ON THE VALUE OF THE INDEX

104. Annual Staff Time Lost By
Staff Vacancy.

EFFECTIVENESS

105. No. Of Staff Hired as Specified
in Job Descriptions

105. & 106.

Proposed index of effectiveness in staffing component.
Comparisons should be made considering scarcity of qualified
personnel.

106. Length of Time Taken to Hire
Staff Specified in Job Description.

EFFICIENCY

107. $\frac{\text{Recruitment \$}}{\text{No. of Specified Staff Hired}}$

107. Proposed index for measuring the efficiency of the staffing
process.

108. $\frac{\text{Annual Lost Staff Time (Weeks)}}{\text{No. of Specified Staff Hired}}$

108. Proposed index for measuring average length of time to fill
vacancy.

Inservice Training

The goal of this is to maintain an informed staff, to introduce new
techniques for use, and to change staff attitude when necessary.

EFFORT

109. Budget (\$ Spent on Training)

109.

110. Hours allocated to Training
the Staff.

110.

111. No. Staff Allocated as Trainers

111.

112. No. New Techniques Presented to
Staff.

112. & 113.

Proposed indices for effort measure related to achieving the goals of inservice training. As stated before, the goals of inservice training are upgrading staff skills, and changing staff attitudes by introduction of new techniques and content lectures.

113. No. Attitudinal Change Topics

EFFECTIVENESS

114. No. Of Staff Who Use Technique
No. Of Staff Who Could Use Technique

114. Proposed index for measuring effectiveness of inservice training in technical change.

115. No. Of Successful Closures After Training
No. Of Successful Closures Before Training

115. Proposed index for measuring effectiveness of inservice training in attitudinal change or technical change. Note should be made that all other conditions must be held constant.

EFFICIENCY

116. No. 26 Closures After Training
No. 26 Closures Before Training
: \$ Spent on Training

116. Proposed index.

COMPENDIUM OF MEASURES

INDEX	COMMENTS ON THE VALUE OF THE INDEX
<i>QUALITY</i>	
117. Experience & Qualification of Training Staff.	117.
118. Instructor Rating Forms	118.
Fiscal Administration	
<i>EFFORT</i>	
119. Knowledge of No. of Sources of Funds Available	119. Proposal for measuring the effort expended in procuring funds.
120. \$ Spent on Fund Raising	120. Proposal for measuring the effort expended in raising funds (appropriate only for private organizations).
121. Budget for Fiscal Administration	121. Proposed index—should include fund raising, accounting, computer time, cost, etc.
<i>EFFECTIVENESS</i>	
122. <u>No. Of Sources of Funds Received</u> No. Of Available Sources	122. Proposal for measuring the effectiveness of procuring funds.

123. $\frac{\text{Budget Received}}{\text{Monies Available}}$

124. Private Donations Received

EFFICIENCY

125. $\frac{\text{Private Donations Received}}{\text{Money Spent on Fund Raising}}$

126. $\frac{\text{Budget for Fiscal Administration}}{\text{Total V.R. Budget}}$

QUALITY

127. Adherence to Regulations as Set Out
by Federal, State, and Local Authorities.

WEIGHTED CLOSURES

128. Rehabilitation Difficulty Score
(Kunce)(56)

129. Difficulty Index Scale
(Miller & Barillis)(88)

123. Proposal for measuring the effectiveness of procuring funds. There is a distinct problem in estimating monies available.

124. Proposal for measuring the effectiveness of fund raising.

125. Proposal for measuring the efficiency of the fund raising operation.

126. Proposed index.

127. Specific to Fiscal Administration (Receipts, Disbursements, Record Keeping, etc.)

Only through the use of a weighted index closure system can a fair evaluation of a program, service delivery, counselor, evaluator, trainer or placement officer be done.

128. Experimental, intended as a screening device to single out more difficult cases. Utility as a closure device must be demonstrated.

129. Experimental, testing found that accuracy was the same as using the Status 26 closures. Cumbersome to use.

COMPENDIUM OF MEASURES

INDEX

- 130. Weighted 26 Closures
(Sermon)(66)
- 131. Weighted Case Closures
(Noble)(75)
- 132. Weighted Closures
(Berkowitz)(74)
- 133. Weighted Closures
(Okla.)(78)
- 134. Florida Difficulty Index
(Fla.)(34)

CASE FLOW

COMMENTS ON THE VALUE OF THE INDEX

- 130. Experimental—currently incorporated in weighted closure systems in New Jersey, Florida, and Minnesota.
- 131. Under development—is directed only to the client who is employed.
- 132. Under development—purpose to develop relative weights for significant client variables. Reliability and validity not established because of often missing client data from case records.
- 133. Under development—first draft of scale seems to be cumbersome, hard to estimate case difficulty of mental disabilities.
- 134. Under development—four basic elements: time in Status 10-26, average case service cost, average number of case services for 26 closures and Sermon's Difficulty index. Florida Index closely agrees with counselor estimate of case difficulty.

All of these indices are valuable for *monitoring* service delivery and any deviations from accepted standards indicates potential problems. There are more useful indices measuring effort, effectiveness, efficiency as detailed in the preceeding sections.

MONITORING

Client Status Rates

- | | |
|---|---|
| 135. % of Caseload in Differing Statuses | 135. |
| 136. % of Change From One Status to Another | 136. |
| 137. $\frac{\text{Total Caseload}}{\text{No. Of Counseling Staff}}$ | 137. Because this index gives an indication of average use of staff time, it is considered important to the "non-informed, significant others". |
| 138. Utilization = $\frac{\text{Use of Service}}{\text{Capacity of Service}}$ | 138. The utilization rate is a useful monitoring measure for indicating the awareness of the need, and consequent use of a service. |

Case Velocity

(14,17,46,56,57,63,109)

- | | |
|--|---|
| 139. Time from—
a. Referral to Acceptance
b. Referred to Placement
c. Acceptance to Closure, Etc. | 139.-141.
Of particular interest are average numbers of clients in pre-service status (10,12) and service interrupted.(24) This index indicates delays in case flow which can result in client dropout, ineffective services, etc. However, high case velocity (speed) can be associated with inadequate services. |
| 140. No. Of Months in a Status
(Client Case Control) | |

COMPENDIUM OF MEASURES

INDEX

COMMENTS ON THE VALUE OF THE INDEX

141. Average Length of Plan

142. Time From End of Plan to
PlacementAttrition
(Drop out rates) (21,118)143. $\frac{\text{Status 10}}{\text{Status 08} + 10}$ 144. $\frac{\text{No. Status 26}}{\text{No. Status 00}}$ 145. No. Status 08
No. Status 26
No. Status 28
No. Status 30146. $\frac{\text{No. Status 28}}{\text{No. Status 26}}$

142. The amount of time should be minimal, however, this may be influenced by local economic conditions, accessibility of employment, or training of the client.

143. Acceptance rate, will reflect backlog, sometimes stated: $\frac{\text{Acceptances}}{\text{All Referrals}}$ 144. Sometimes stated as $\frac{\text{Closed Rehabs}}{\text{Referral}}$ or $\frac{\text{Total Status 26}}{\text{Total Closures}}$ (also measure
(Status 08,26,
28,30) of effectiveness)

145. These client counts, in and by themselves are meaningless, unless they are part of a ratio or rate. (Service Efficiency).

146. May indicate client hostility to program.

V EXAMPLES OF PROGRAM EVALUATIONS

Sixteen abstracts have been prepared to provide the reader with practical illustrations of program evaluation plans, completed program evaluations, systems analysis and cost benefit studies. Each represents a different approach to the statement of the program-evaluation problem, operationalization of evaluation goals, methodology, collection of data, analysis of the data, interpretation of the data and, finally, the conclusions. In short, if the evaluator reads one of these evaluation final reports, he would have a model to follow in conducting "his" program evaluation. Final reports may be secured from the study's author, from the National Technical Information Service (NTIS, Springfield, Va. 22151) the cited journal, Social and Rehabilitation Services, DHEW (330 C Street, S.W., Washington, D.C. 20201) or from the RUL Information Center, ICD Research & Rehabilitation Center (340 East 24th St., N.Y., N.Y. 10010).

PROGRAM EVALUATION PLANS

HERZOG, ALLEN: *A MODEL FOR THE EVALUATION OF AN ACTION-RESEARCH PROGRAM*. NOVA SCOTIA NEWSTART CORPORATION, YARMOUTH, NOVA SCOTIA, CANADA, MARCH 1970.

Summary

This paper discusses a social action program-evaluation methodology based on experience of the Canadian NewStart Action-Research (i.e., demonstration) Project. NewStart is described as an effort to approach problems of socioeconomic disadvantage in a systematic, scientific manner.

The paper presents a plea for a conceptual model for appropriate evaluation designs to be built into demonstration projects. Its stated (and achieved) purpose is adequate conceptualization of the nature and problems of action-research, with particular respect to program objectives, evaluation requirements, and program impact or outcomes. The model includes a functional classification of action-research intervention strategies (i.e., educational, facilitative, and structural) and evaluation considerations particular to each interventive strategy.

Two main types of action-research evaluation are clarified as process evaluation and outcome evaluation:

- 1) Process evaluation is concerned with program monitoring and the feedback of evaluative material for program improvement.
- 2) The outcome approach focuses instead on social-action programs. Five types of outcomes are considered:
 - Policy information evaluation
 - Product evaluation
 - Research evaluation
 - Behavioral change evaluation
 - Service evaluation

Evaluation considerations are discussed with respect to each type of output goal. In addition, inputs for all the output goals are specified.

The paper concludes with a consideration of statistical methodology in action-research evaluation, and barriers to the attainment of action-research outcome objectives.

Comments

A valuable model providing concrete assistance in clarifying the types of data that are obtained from varying evaluation approaches and the applicability of these data. Has utility in guiding systematic evaluation design, and in clarifying program and evaluation interfaces and relationships. A useful conceptual plan-

ning tool with "substance" based on a sound understanding of evaluation methodology. It is "systematic" in the sense that it considers relationships between program strategy and research methodology, as well as between evaluation inputs and outputs. It is a sophisticated design.

MICHIGAN STATE: *EVALUATION PLAN FOR THE DEMONSTRATION PROJECT COORDINATING WELFARE AND REHABILITATION FOR THE DISABLED PUBLIC ASSISTANT RECIPIENT*. DEPARTMENT OF SOCIAL SERVICES, LANSING, MICHIGAN, 1971, director, B. HOUSTON (SRS RD 11-P-5720 4/5-01)

Summary

The focus of the evaluation plan is upon measures of efficiency as they relate to program effort and effectiveness. Two program objectives are stated to guide the plan.

- 1) Vocational rehabilitation services to Aid for Disabled recipients are to be more efficiently delivered.
- 2) The coordination of Vocational Rehabilitation Services and Department of Social Services systems is to be tested as a means for increasing the efficiency of service delivery.

There are separate evaluation designs for the two objectives:

- 1) A survey research approach comparing base-line data from 1971-72 with later project data, to establish the efficiency of the delivery of services to AD recipients; and
- 2) A modified systems analysis to test the concept of coordination as a means of increasing service delivery efficiency.

In order to meet the first objective, efficiency of operation of the PA/VR program, four types of information will be collected and analyzed:

- 1) Descriptive client information.
- 2) Measures of program effort, such as services rendered to each client, amount of time vocational counselors and service workers spend together, and length of time between acceptance of each case to closure.
- 3) Measures of program effectiveness, such as decreased time from application to service plan will be compared with successful rehabilitations to determine whether the PA/VR program increased earned income per AD recipient.
- 4) Measures of program efficiency, such as the determination of other reha-

bilitative services for Aid to Disabled recipients in the community, the cost of these, and calculation of the cost of removing barriers to employment for AD recipients in the demonstration project. The data used for the variables or indices is, for the most part, routinely collected VR data.

In order to meet the second objective concerning the efficiency of inter-agency coordination, a framework for analysis was developed. This is composed of two components:

- 1) A method for identifying and measuring interagency coordination by developing indices of administrative cooperation and of cooperation between workers of the two agencies.
- 2) The second component is a framework for explaining the coordination by observing the nature of the essential contributing factors. This includes characteristics of the organization, characteristics of the organization's environment, interorganizational factors and the agencies' view of the public interest.

Data will be collected from staff meetings, interviews, memos, service plans, progress forms and forms to identify joint decision-making between VR counselors and SS workers.

Comments

This is a strong design that is feasible to perform at a central vocational rehabilitation office. Clearly, the design seeks to evaluate the objectives of the National PA/VR program. Implicit within this approach is an evaluation of some of the aspects of program management, an overall evaluation of service delivery and an evaluation of a few of the factors in counselor performance.

OHIO STATE UNIVERSITY: *PLAN FOR EVALUATING THE PA/VR FRANKLIN COUNTY EXPANSION GRANTS PROJECT*. SCHOOL OF SOCIAL WORK, OHIO STATE UNIVERSITY, COLUMBUS, OHIO, OCTOBER 1972, author, L. SCHNEIDERMAN, (SRS, RD. 25-P-65-439/5-01)

Summary

The focus of the evaluation design is upon measures of program effectiveness and measures concerning the potential utility of rehabilitation technology for the public assistance population. Four evaluation questions are asked in guiding the design:

- 1) What benefits and costs are associated with PA/VR Expansion Grant activities?

- 2) What differences in service outcome occur for the sub-population served by the project?
- 3) What is the relative size of each of these sub-population groups within the total population?
- 4) What patterns of interagency team work appear most effective in service delivery and what training program might most effectively prepare staff for such practice?

This program evaluation plan is a longitudinal research design utilizing random samples of 500 clients under care compared to a matched sample of clients without care.

Thirty-four specific research questions are asked. Primarily these are directed to the general areas specified in the evaluation questions. That is:

- 1) Which services provide which client populations with the most successful closures?
- 2) What is the relative position of the agency within the community?
- 3) What are staff capabilities?

Data will be collected during five interviews: the initial interview, a closure interview and three follow-up interviews at three months, six months and one year. The measuring instruments used to collect information will be the R-300 and supplements such as a face sheet for client ID purposes, the Psychological Well-Being Index, a specially designed Job-Satisfaction and Family-Functioning Index utilizing Cantril's self-anchoring scaling technique, the Health Index from the National Health Survey, the Alameda County, California, Health Ways of Living Survey and the Community Survey of Aging.

Comments

This is a strong research design for program effectiveness in the general area of behavior changes. It will provide basic information concerning what services are effective with what client populations. At the writing of this draft plan, no provision is made for making comparisons concerning the efficiency or quality of the program's operation, or the quality, efficiency, or effectiveness of counselor performance.

Caution is offered to the new evaluator. If he should elect to replicate this design, it is sophisticated, time consuming, will require many man hours to complete and therefore will be expensive to reproduce.

UNIVERSITY OF WISCONSIN: *AN EXPANDED PROGRAM OF VOCATIONAL REHABILITATION METHODOLOGY AND DESCRIPTION OF CLIENT POPULATION*, R.R.R.I., UNIVERSITY OF WISCONSIN, MADISON, WISCONSIN, 1970, author, G.N. WRIGHT. (SRS 12 55208, NTIS PB 210 265)

Summary

The purpose of the Wood County Project is to initiate a model community rehabilitation program lasting five years with intensive rehabilitation services for a physically and culturally disabled client population.

There are four objectives in the long-term planning for the project:

- 1) Increased effectiveness and efficiency of a bigger and more comprehensive program to provide services to a wider scope of clients.
- 2) Development of administrative guidelines for the transition of the program's operation during the expansion process.
- 3) Testing of theories and techniques applicable to the professional function of the rehabilitation counselor.
- 4) Development of appropriate instrumentation for evaluating the experimental program.

The evaluation study uses a social research design comparing a traditional vocational rehabilitation agency to the "expanded experimental" program.

The research questions center on the assessment of the experimental programs' impact on clients, on change in intra-agency processes, and on community.

Twelve different instruments (scales) are used to measure change in clients, intra-agency process and community impact. Of the twelve, five are standard scales and the other seven were developed for this study. Data are collected from:

- Wisconsin's Data Record form
- Client service records
- DVR Fiscal tapes
- Rehabilitation counselor master list of clients
- Personal instructions with professionals

Comments

This is a comprehensive program evaluation plan that considers two measures of effective program outcome, (1) change in clients' behavior and (2) the community's attitude toward the program. The community's attitude toward the program is based on the economic feasibility of the program and the extent of public knowledge concerning the program.

The assessment of the experimental programs' impact for changing the intra-

agency processes is not as well defined. This problem centers on the definition of intra-agency processes. Clearly considered, intra-agency processes are the selection of clients and case management. This only represents decision-making at the lowest authority level and does not reflect management policy.

RESEARCH APPROACH

NORTHEASTERN UNIVERSITY: *STRUCTURE AND DYNAMICS OF SOCIAL INTERVENTION*. MONO. 9, RRRI, D.C. HEATH AND CO., LEXINGTON, MASSACHUSETTS, 1970, author, G. SPENCER (SRS RD 3017G, NTIS PB 216 320)

Summary

This is a longitudinal study comparing the effectiveness of three types of organizational structures (bureaucratic, community action, and community service) in vocational rehabilitation programs.

The objectives of the study are to identify the structural and dynamic factors which cause the three projects to function the way they did.

Seven structural factors of the program environment are identified:

- Funding agency
- Parent organization to which funds were awarded
- Project director
- Initial staff
- Existing social welfare system
- Potential users of the agency services
- Wider community

The six organizational phases defined in the instruction process of the projects are:

- Obtaining initial resources for the program
- Initial organization of the project
- Struggle for system integration
- Conflict resolution
- Reshaping the organization
- Development of a relatively stable structure and process

Specific comparisons were made to establish:

- How each program went about operationalizing its goals and so developed structure and process

- Why each program developed as it did
- Measurements for the varying effectiveness of client outcome under the three projects

Data are screened from formal documents (initial project proposals, annual progress reports, memos, letters, etc.) observations, communication, interviews and client files.

Comments

The effectiveness study included is a strong evaluation design. Some of the indices used are rehabilitated, full or part-time work, skilled or unskilled work, salary, welfare and demographic comparisons.

While this is one of the few studies examining the effectiveness of differing organizational milieus (structure units and functional processes) the analysis is "soft". That is, no indices or tools for quantifying the assessment of an organization's developmental outcome are shown. However, a conceptual model is presented that can guide a program planner, evaluator or administrator to be aware of the various system units and processes. The reader is alerted to the necessity for sensitivity to structural and functional process interaction and their impact on program outcome.

If this concept were "operationalized", it would provide a powerful instrument for the heretofore neglected comparison of program-organization influence on program effectiveness. Any evaluation which encompasses several different projects should consider this conceptual model.

UNIVERSITY OF MISSOURI: *REHABILITATION AND THE CULTURALLY DISADVANTAGED*. MONO. 1, RRRI, COLUMBIA, MISSOURI, 1969, editor, J.T. KUNCE AND C.S. COPE. (SRS RD 2326G, NTIS PB 210 219)

Summary

This is a study conducted by searching the literature, methodologically defining issues from that literature and then drawing conclusions. The issues defined surround the topics of characteristics of disadvantaged clients, counseling, program outcome (effectiveness) and program organizational structure. This study is included here because it clearly presents some major field issues with their contradictory findings. The resolution or qualified resolution of these issues will only come through intelligent program planning design and repeated evaluations in a true field setting. With this information then available, the field can mature and allow the evolution of more efficacious programs.

Of particular interest is the survey of traditional counseling concepts and their applications. Complementary and contradictory study findings are noted and summarized. In addition, counseling practice issues are defined and dis-

cussed. Another topic of interest is the definition of program's primary approach (change behavior, change socially, promote jobs, provide jobs, etc.) and a discussion of the effectiveness of each approach. The final discussion investigates the problems surrounding the present "Bureaucratic-Professional Model" of organizational structure.

Comments

This is not an easy research project. However, clear discussion of issues surrounding the topics that concern program effectiveness and efficiency is useful to the new evaluator. Further, the issues are defined in researchable form. In other words, the reader can readily understand if he is evaluating his program for effectiveness, what the program objective should be, that is, to change the client or society. While the title leads the reader to assume the topic is specific to the culturally disadvantaged, it defines universal issues with enormous sources of back-up data pro and con.

UNIVERSITY OF MINNESOTA: *A FOLLOW-UP SURVEY OF FORMER CLIENTS OF THE MINNESOTA DIVISION OF VOCATIONAL REHABILITATION*. MINNESOTA STUDIES IN VOCATIONAL REHABILITATION XXVI, MINNEAPOLIS, MINNESOTA, APRIL 1969, author, H.E.A. TINSELY; R.G. WARREN; D.J. WEISS. (SRS RD1613.12)

Summary

The Minnesota follow-up survey reports on the effectiveness of the Vocational Rehabilitation process in the state of Minnesota from 1963 to 1967. It compares the closed-placed rate of rehabilitated clients to the employed-at-acceptance rate. In addition, it compares the rate of employment of a small group of "not rehabilitated" clients to a group of "rehabilitated" clients.

Specific questions asked:

- Do clients who are rehabilitated by DVR stay employed?
- Are they satisfied with their jobs?
- Are they satisfactory workers?
- Are their earnings competitive with other workers?
- What kinds of jobs do they obtain?
- Do they work full time?
- Do they become self-supporting?

Data were collected from a mail questionnaire of 6435 clients. Post card reminders and telephone inquiries were used to follow-up on the original questionnaire. There were 3977 usable responses.

Comments

This survey which is simple in design, provides a method for establishing base-line data for state standard setting or future effectiveness studies for that state. It also provides a basic model for a simple effectiveness study. As simple as this survey appears, a word of caution must be offered: follow-up on 6000 clients seen over a four-year period represents a herculean undertaking. An alternative would be to use selected samples. This survey with the addition of more comparisons between treatment vs. no-treatment groups would be an excellent example for an evaluator's retrospective-effectiveness study.

WASHINGTON: *EVALUATION AND INVENTORY PROJECT OF PUBLIC ASSISTANCE RECIPIENTS*. VOCATIONAL REHABILITATION DEPARTMENT OF SOCIAL SERVICE, OLYMPIA, WASHINGTON, 1972, author, R. SCHUMANN (SRS RD 25P65216/0-03)

Summary

The objectives of this Public Assistance and Vocational Rehabilitation Project include:

- In depth screening of Public Assistance recipients which will identify those recipients with rehabilitation potential
- Provision of accurate information concerning the need of Vocational Rehabilitation services
- Testing the feasibility of the Team Evaluation concept for Vocational Rehabilitation
- Expansion of the quantity of services provided to Public Assistance recipients

The hypotheses tested in this program evaluation are:

- Within the group of Public Assistance clients traditionally considered unemployable, there is a substantial group who can benefit from rehabilitation services
- The use of a multi-disciplinary team to determine rehabilitation potential is more effective than individual caseworker selection and subsequent referral to Vocational Rehabilitation
- Hard-core, unemployable Public Assistance clients have a variety of unmet service needs
- Certain personal characteristics can be isolated and shown as having a relationship to rehabilitation potential
- Social and economic benefit which will accrue as the result of clients' participation in the Project will outweigh the costs of the Project

- Clients who are accepted by the Project and who subsequently drop from the Project or who complete the Project, but are unable to maintain steady employment, will experience social gains
- When resources which a client needs to complete his rehabilitation plan are not available within the community, these resources can be developed or made available through community organization activity

The project uses the following innovations in the delivery of client services: the team approach, a policy of "aggressively reaching out" to clients, joint structured interviews, and multi-integrated services.

The methodology used to evaluate this program is to compare 80 clients who receive project services to 79 clients, who though eligible for project services, receive traditional community services. This is done in four substudies: a description of the study's population, an examination of the service delivery process (client evaluation, and plan implementation and case closure), an effectiveness (follow-up) study and a Cost/Benefit study. Data are collected from social summaries, a structured interview, pre-screening data from Heimler Scale of Social functioning and case histories.

Comments

This evaluation is divided into four substudies which address themselves to the hypothesis. The information for the hypothesis is compiled, with the exception of the Heimler Scale, from the usual forms and records.

The study represents a feasible example of a program evaluation for the new evaluator or uninitiated project director. However, while it is presented in a very readable manner, some extra time must be spent in reconstructing the methodology for replication of the step-by-step process.

SYSTEMS APPROACH

NORTH CAROLINA: *PROGRAM EVALUATION*. NORTH CAROLINA DIVISION OF VOCATIONAL REHABILITATION. Author ROBERT PHILBANK, ASSISTANT DIRECTOR FOR PLANNING AND EVALUATION.

Summary

There are basically three management systems utilized (or currently being developed) by this Agency to ascertain 1) effectiveness, 2) efficiency, and 3) cost/benefit of the various programs. These are:

- A. Advanced Casework Management System (ACMS) (casework level). This system is currently being developed. It will utilize general systems theories

and seeks to develop a more scientific basis for evaluation and programming for clients. A part of the system is a rehabilitation plan (or program) which consists of an organized, comprehensive set of behavioral objectives to be accomplished within a given time frame with prescribed services, scheduled evaluation toward accomplishment of behavioral objectives and cost estimates for attaining each objective. When such a system is developed, the agency will better be able to define those factors which are significant to vocational participation, sources of potential gains which clients (or groups of clients) can make and the relative effectiveness and cost of individual modalities in attainment of the objective. Thus, a weighted case-closure index will become a part of the ACMS, similar to that developed by Duane Sermon of the Minnesota DVR.

- B. Operational Management System (OMS) (program level). This is a short-range planning system which is based on the concepts of: 1) management by objectives (MBO) and, 2) planning, programming, and budgeting (PPB). The rehabilitation services provided by this Agency are divided into seven major program areas. These include behavioral disorders, mental illness, developmental disabilities, physical disabilities, communicative disorders, disabled disadvantaged (PA/VR), and S.S.D.I. Thus, one of the major objectives of the OMS is to define program structures. The items which are defined and analyzed by the OMS are: the target population of the respective programs, the program goals, the objectives of the program in terms of the numbers to be evaluated, served and rehabilitated, and the resources (case-service monies) that will be required to accomplish the goals and objectives.

Data concerning the accomplishments (numbers served) and the allotments (case service funds) are fed into the computer as base-line data, and the efficiency, effectiveness, and costs of each program is measured monthly. Formal reviews are held quarterly. Examples of data generated are:

- 1) Program Effectiveness—The relationship between the predetermined intent of the program (objective) and the actual program outcome (accomplishment). For example, if the intent of the program is to rehabilitate 100 persons during a given time period, and it actually accomplishes 75, then it is considered to be 75% effective.
- 2) Program Efficiency—The relationship between the resources consumed (costs) and the program accomplishments (output). To perform this analysis, the cost per output unit (rehabilitant) for each program is determined and compared with programs of the same type. That is, one PA/VR program is compared with another PA/VR program, but a PA/VR program is not compared with a physical disabilities program.
- 3) Cost/Benefit—This concept refers to the relationship between resources

consumed (costs) and value accrued (benefit). The time frame is one year and the formula is:

Benefits =

$$\frac{(\text{Income at Acceptance} - \text{Income at Closure}) + \text{Reduction in P.A.}}{\text{Total Costs}} / \text{Year}$$

- C. Developmental Management System (DMS)—This system will be designed to allow the Agency to do advanced planning based on Regional needs and to measure the effectiveness of operational programs as they operate independently and in concert in meeting total service population group needs.

Comments

This is a state plan for the on-going evaluation of vocational rehabilitation. Two program components are evaluated: 1) casework management and 2) general management. This plan presents a way to catalogue program components and then analyze them. The data are grouped into the popularly discussed variables, measures, and scales to measure program effectiveness and efficiency.

This is a very viable evaluation plan for a centralized Vocational Rehabilitation Office if central management uses a systems orientation to the general management process and the Planned Program Budgeting System (PPBS) is used for the budgeting process.

UNIVERSITY OF DENVER: *A CONCEPTUAL FRAMEWORK FOR THE DESCRIPTION AND ANALYSIS OF HUMAN SERVICE DELIVERY SYSTEMS*, SOCIAL WELFARE RESEARCH INSTITUTE, DENVER, COLORADO, 1972, authors, A. SELIG, R.W. GAGE, P. LANDON, C. GOERDEL, E.C. BAUMHEIER. (SRS RI 5602-C1)

Summary

The purposes of a conceptual framework is to develop practical instrument for use in the study and analysis of human services. The need for such an instrument has repeatedly been demonstrated when attempts to concretely describe the nature and operation of the human service delivery systems have been made.

The authors express the opinion that programs should be analyzed on three interrelated levels:

- System variables: structure, function and process
- Performance: maintenance, efficiency and output
- Program impact: outcome, input and effectiveness

However, this conceptual framework concerns only the system variables. An outline & definitions for the interrelationships of these variables is presented in the study on page 00. Briefly, the conceptual framework may best be visualized as a matrix. On one axis the *structural* elements (Policy making, management, and operations) of a system are listed. On the other axis the *functions* (problem identification, policy formulation and goal setting, planning program operation and evaluation) of a system are listed. The program's process describes the activity that occurs when a structural element is required to perform its function.

In order to use this conceptual framework, a series of questions have been prepared so the program evaluator may repeatedly request standard information.

Comments

The authors stress that the model is preliminary; validation has yet to be verified. However, it presents to the reader one scheme for examining structure, function and process within human service delivery systems. It would be difficult to apply this model without some knowledge on the part of the evaluator of systems theory, and without seeing a practical application. Future users should consult the authors for examples of practical applications.

OREGON: A PROGRAM AND EVALUATION SYSTEM MODEL FOR REHABILITATION AGENCIES: VOCATIONAL REHABILITATION DIVISION, EUGENE, OREGON, 1971 (SRS RD 2987 & 12 55 297, NTIS PB 215 431)

Summary

This report is based on a two-year (1969-1970) project undertaken by the Oregon Department of Vocational Rehabilitation to adapt business-oriented management-by-Objectives (PIE-PPBS) to the needs of a state rehabilitation agency. The report covers the conceptual approach and general methodology of the project, and notes that extensive project "technical blueprint" material is available on request.

The project was a comprehensive and integrated systems effort to develop a means for collecting adequate and relevant information organized in a manner to lead to improved vocational rehabilitation program and policy decision-making. Its purpose was to provide a rational basis for program resource allocation. It included data for realistic forecasting and program evaluation according to defined measurable goals, objectives, and standards, including output or impact measures.

Thirty-two analytical tools (i.e., measurements) were developed, tested and revised during the course of the project. Development of the analytical tools was based on fifteen major studies in four content areas: (1) refining goals, objec-

tives, priorities and standards; (2) identifying problems; (3) organization of problem-solving elements and alternatives; (4) use of resources. Examples of the analytical tools include: resource inventory; data inventory; client data recording system; forecasting model; budget allocation model.

The project developed its information tools in a stepwise fashion. Program elements were each tested, interfaced and implemented as soon as each was developed and reviewed. Thus, the total management information system was developed into an operational network incrementally out of project developed and tested elements, rather than being designed in toto before the parts were tested or operationalized.

The project report includes operational recommendations for other agencies interested in applying a similar information systems approach to their own shops.

Comments

A report of a functioning service impact-oriented, comprehensive, management-by-objectives information system that has a practical and specified methodology. The project represents an extensive allocation of resources and a level of operationalization that should save other organizations interested in the management-by-objectives approach significant time and resources by borrowing from this project's conceptual framework, methodology, and developed information measurement tools. Represents the kind of effort needed to provide impact measures related to resource allocation and program and policy alternatives. Should have widespread applicability—a sound foundation for other agencies to borrow from and build upon.

AUERBACH CORP. *WIN SYSTEMS ANALYSIS: FINAL REPORT AND WIN MODEL*. PHILADELPHIA, PA., 1971, author, M. SATINSKY (NTIS P B 199 101)

Summary

This report has two major components: (1) a descriptive analysis of the system under which the Work Incentive Training Program (WIN) is administered; and (2) major recommendations for improving the WIN program. The latter was developed as a model WIN system and based on the descriptive analysis of the existing system.

The descriptive analysis of the operational WIN system is based on interviews with WIN administrators and line staff and data from reviewing WIN enabling legislation, planning and program documents, agency and state guidelines, program issuances and policy memoranda. These data were gathered from all administrative levels (national, regional, state, and local project sites). The descriptive analysis has three major component parts. These include analysis of the con-

ceptual and legal framework for WIN, analysis of the WIN management system, and analysis of local operations and how they are affected by the program assumptions, legislation, and management system.

The model WIN system has the same three major component parts. For each component part, the model format includes a summary of major problem areas, recommendations for changes, and the rationale for the recommendations. The model itself is guided by the consideration of developing an "ideal" system within reasonable constraints and limitations.

Comments

This type of descriptive assessment provides a valuable function in a systems approach to program evaluation by systematically exploring the program premises, conceptual background and legislative directives which form the backdrop out of which and against which operational programs are developed at the local level. Its focus is descriptive explication of often overlooked fundamental program inputs and the parameters that are set for program outputs—an approach that might be employed to determine the basic structural reach and limitations of any social intervention program. A policy analysis approach based on conceptual analysis, narrative documentation and some supporting statistical data.

COST BENEFIT

WASHINGTON: *COST/BENEFIT STUDY, FINAL REPORT: EVALUATION AND INVENTORY PROJECT OF PUBLIC ASSISTANCE RECIPIENTS*. VOCATIONAL REHABILITATION DEPT. OF SOCIAL SERVICES, OLYMPIA, WASHINGTON, 1972, author, R.A. SCHUMAN (RD 25-P-65216/0-03)

Summary

As one component of this project, a Cost/Benefit analysis of selected VR clients was used as one tool in an assessment of the rehabilitation potential of public assistance clients who had traditionally been considered unemployable.

Comments

Cost/Benefit measures used were simple aggregative ones readily available or computable from administrative records, and the analysis applied to the measures was a simple rather than sophisticated comparative process. Thus, the approach utilized here should be well within the capacity of most agencies. However, the power of the evaluation is reduced by the simplicity and aggregate

nature of the measures and analysis used; thus, it should *be used as one indicator of rehabilitation potential along with a battery of other indicators*. A seemingly feasible approach which should provide a fairly gross indicator rather than a refined analysis.

BORUS, M.E.: *A BENEFIT/COST ANALYSIS OF THE ECONOMIC EFFECTIVENESS OF RETRAINING THE UNEMPLOYED*, YALE ECONOMIC ESSAYS, YALE UNIVERSITY PRESS, NEW HAVEN, CONNECTICUT, Fall 1964

Summary

The purpose of the Study is to examine the benefits and costs of alternative manpower training programs (including those under MDTA) from the perspective of the individual, the government, and the economy.

The analytical boundaries are limited to the quantifiable economic variables.

A sample of 373 workers (312 men and 61 women) was drawn from Connecticut retraining programs. The group of workers who successfully completed the retraining course and utilized the retraining is compared with the groups of workers who (a) completed the training course, but did not utilize the skills learned, (b) withdrew from retraining without employment, (c) refused retraining without employment.

The measures involved and analytical techniques used determine the net economic benefits of retraining to the individual worker who utilized the retraining skills by calculating (a) the increased wage income and (b) the reduced unemployment, simultaneously calculating the off-setting factors such as: (1) increased taxes, (2) decreased unemployment benefits, (3) occupations of the retrainees in the absence of retraining, (4) occupational mobility out of retraining-related occupations, and (5) a discount for time preference.

The existence of such benefits for the workers in the sample was found by using multiple regression techniques and by inspection of actual employment records. Also, the effects of the differences in demographic characteristics on the employment records of the four groups are treated in intergroup comparisons by using multiple regression techniques.

The findings were that benefits far outweighed the costs in each of the three cases and in the following descending order:

- (1)economy (accounts for the highest figure), then
- (2)the government, followed by
- (3)the individual worker.

Comments and Critique

Although extensive regressions were run to determine the effects of the retraining program on individual earnings, only partial results are presented. Thus, for example, although race was included as one of the independent variables, its significance in explaining income variations is not reported. In many instances the regression results are not highly significant statistically (i.e., Low "t" ratios). The author has sought to demonstrate in the regression analysis that retraining programs do lead to increased earnings, even after controlling for such factors as age, education, labor force attachment and participation, previous training, income, unemployment, and unemployment benefits.

The major contribution of the study is the evaluation of retraining programs from three perspectives. This successfully separates out the differences in impact on the individual, the economy, and the government.

While it has not considered any of the noneconomic factors, it has included and examined thoroughly all of the economic variables.

The sample included a small percentage of women workers, but even this small sample is finally omitted in the analytical calculations. The research is limited to an exclusively male sample study.

The findings derived from a sample study of the Connecticut area cannot be generalized as applicable to the entire nation, especially when the living indices differ so much from state to state.

CONOLEY, R.W.: *A BENEFIT-COST ANALYSIS OF THE VOCATIONAL REHABILITATION PROGRAM*. JOURNAL OF HUMAN RESOURCES, VOL. 4, SPRING 1969.

Summary

The purpose of the Study is to re-evaluate through CB analysis vocational rehabilitation programs already evaluated through 1967, using some refinements in the methodological techniques and adjusting for the earlier results to allow for the current changes in tax legislation.

The analytical boundaries name some of the social variables that ought to be considered in the study, but actually use only the quantifiable economic variables.

The sample is drawn from client data from State-Federal vocational rehabilitation programs between the years 1958 to 1967.

The measures and analytical techniques used estimate C/B ratios for vocational rehabilitation programs. Benefits are computed by taking into consideration the following items:

- (1) Discounted value of increased lifetime earnings after rehabilitation programs

- (2) Tax revenues generated by the employed rehabilitants
- (3) Savings in government transfer payments.

The major components of program costs are adjusted to take account of costs incurred outside the program as well as loss of earnings incurred while in the program.

The findings conclude that VR programs are highly successful.

However, wide variations in C/B ratios are noticeable in individual rehabilitants. One finding states that rehabilitants with highest earnings at closures are also the ones with the highest earnings at acceptance *and* are the most expensive to rehabilitate. Therefore, the VR program is economically more efficient to the less productive disabled.

Comments

This study, like all others of its kind, finds that VR programs are highly beneficial.

The generic weakness of this kind of study is the choice of discount rate. Needless to say, if the discount rate chosen is sufficiently low, any program could be made to look profitable. The discount rates chosen without *a priori* justification are 4 percent and 8 percent.

A second major problem is that with most CB/CE studies, while giving recognition to social variables that probably have significant impacts on CB/CE factors, only easily quantified economic variables are actually used.

GRIGG, C.M. ET AL: *VOCATIONAL REHABILITATION OF DISABLED PA CLIENTS: AN EVALUATION OF FOURTEEN RESEARCH AND DEMONSTRATION PROJECTS*. NO. 8 INSTITUTE FOR SOCIAL RESEARCH, FLORIDA STATE UNIVERSITY, TALLAHASSEE, 1969. (SRS RD 1323 G, NTIS PB 208 446)

Summary

The purpose of the study is to evaluate vocational rehabilitation research and demonstration projects through an economic analysis of the cost and benefits of rehabilitating disabled public assistance clients.

The analytical boundaries are limited only to the economic variables, for the purpose of determining the resource allocation decisions. At the same time it does recognize the insufficiency of such an analyses in considering the range of possible impacts of the vocational rehabilitation process.

The sample is drawn from data on 1100 clients from fourteen research and demonstration projects.

The measures involved and analytical techniques used on the theoretical

framework are the usual ones of discounting increased lifetime earnings and relating them to cost of programs and transfer payments.

The costs of treatment is regressed on the following dummy variables—sex, race and education. The study estimates the probability of rehabilitation by regressing a dichotomous dependent variable (success/non-success) on a set of characteristics of the individual (a major contribution of the study).

Estimates are computed for four sets of regression equations. The first two regressed costs of individuals who were rehabilitated and those who were not rehabilitated, on age, sex, race, ethnic group, education, type of disability, type of public assistance received at referral and presence or absence of spouse. The third and fourth equations regressed the change in weekly earning and success of rehabilitation, on the same independent variables.

Results from the regression equations are used to estimate the Cost/Benefit (CB) ratio for a subsample of clients—AFDC recipients who had their spouses present. The results are presented in terms of male/female; White/Negro/other; and age groups between 30-50 with education less than eight years. In each case C/B ratios are computed for physical and mental disabilities.

Finally, the study estimates the impact of savings of transfer payments.

Findings claim that investment in VR programs for public assistance clients is a profitable activity. The returns to the public assistance clients of the VR program and to the society which supports the public welfare program are high enough to warrant the continuation of the program. Out of the 1146 clients who were "closed-rehabilitated," only 6 percent remained unemployed after their cases were closed.

It is concluded that it is possible to remove the disabled public assistance clients from the welfare rolls by providing them with rehabilitation services.

Comments

Although the author recognizes the tenuous nature of his regression exercise (he violates the assumption of the classical regression model) he goes on to use the regression results (a great proportion of which are statistically not significant) to estimate benefits and costs. As such, his benefit-cost ratios are, at best, highly tentative. However, he has shown that the subject may be approached with some amount of empirical comprehensiveness. His use of the profit model to estimate the probability of successful rehabilitation is legitimate and highly useful.

OTHER FIELD EXAMPLES

1. California Department of Rehabilitation, *An Individualized Approach to the Problems of Poverty and Dependency*, by Z.L. Gulledge. Sacramento, Calif.: SRS RD 2089, 1970. (NTIS PB 215607)
2. Florida State University Institute of Social Research, *Vocational Rehabilitation of the Disabled Public Assistance Client*, by C.M. Grigg, et al. Tallahassee, Fla.: 1969.
3. Maryland Division of Vocational Rehabilitation, *A Follow-up Study of 1365 referrals for Rehabilitation Services*. Baltimore, Md.: SRS RD 2415-G, 1972. (NTIS PB 208921)
4. Michigan Department of Rehabilitation, *Two Years Later: A Follow-up Study of Cases Closed by the Michigan Vocational Rehabilitation Service in fiscal year 1970*. By R. Struthers. Lansing, Mich., 1972.
5. Northeastern University Regional Rehabilitation Research Institute, *Psychodynamics and Enablement in the Rehabilitation of the Poverty-Bound Client*, by G.J. Goldin, et al. Lexington, Mass.: Heath-Lexington Books, SRS RD 30176G and RD 1621G, 1970.
6. Sewell, D.O., *Training the Poor: A Cost Benefit Analysis of Manpower Programs in the U.S. Anti-Poverty Program*. Kingston, Ont.: Queens' University Industrial Relations Centre, 1971.
7. Smits, S.J. and W.J. Aiken, *Descriptive Study of Supervisors Practices as Perceived by the Counselors in State Vocational Rehabilitation Agencies*. Bloomington, Ind.: SRS RD 2780-G, 1969. (NTIS PB 209 777)
8. Tella, D., et al., *A Model for Manpower Training Evaluation*. McLean, Va.: Planning and Research Corp., 1970 (NTIS PB 200 555)
9. Threlkildson, P. and P. Reno, Cost-benefit evaluation of the Bernalillo County Work Experience Project, *Welfare in Review* 6 (2) (March-April, 1968).
10. University of Arkansas, *Rehabilitation of the Rural Hard Core Welfare Client Utilizing Rehabilitation Center Services*, by L. Dickerson, et al. Fayetteville, Ark.: Rehabilitation Research and Training Center, SRS 12-P-55218, 1972.

11. University of Missouri Regional Rehabilitation Research Institute, *Rehabilitation in the Concrete Jungle*, by J.J. Kunce, *et al.* Columbia, Mo.: SRS RD 2326G, 1969. (NTIS PB 210 291).
12. University of Oklahoma Regional Rehabilitation Research Institute, *Management Practices in Vocational Rehabilitation District Offices*, by W. Hills and H. Vialle. Norman, Ok.: 1973.

GLOSSARY

ARRAY 1) A series of numerical values arranged in classes. 2) A series of values arranged according to magnitude.

BIT (pl. bits) Smallest unit of information used in computer science.

COMPONENT (see *element* used in IRS manual) Program classification subordinate to program level and above program element (e.g. program management, *staffing*, recruitment).

CONTROL VARIABLE (see also *Variable*) An extraneous undesirable that is operating in an experimental situation in addition to the Independent Variable (*q.v.*), and having an effect on the Dependent Variable (*q.v.*). These variables may be held constant or randomized so that their effects are neutralized or cancelled out.

CRITERION (pl. criteria) 1) Standard by which the general validity of a method, measuring device, formula, etc. is judged. 2) Standard of performance on which a judgement or decision may be based. 3) Same as dependent variable.

DATA (s. datum) Something used as a basis for measuring. 2) A collection of statistical facts and figures (usually indicates some processing).

DATA ELEMENT 1) A specific item of information appearing in a set of data. 2) Smallest unit of information to which reference is made.

DATA PROCESSING 1) Operation of converting crude information or Raw Data (*q.v.*) into a usable or storable form. (115) 2) Numerical computation.

DELPHI METHOD OF EVALUATION The process of systematically refining the opinions of experts in evaluating programs. This technique involves: a) obtaining opinions of experts by use of questionnaires, b) controlled sharing (or feedback) and consequent reformulation of questionnaire results with the experts, and c) aggregating individual opinion into an overall group judgement. (46)

DEPENDENT VARIABLE 1) Outcome variable of a study or other activity (e.g. one dependent variable in VR is client employment). 2) That which is the "effect" in a cause and effect relationship.

DESCRIPTIVE STUDY (see also Observational and Experimental Studies)

Describes what happened in a given situation. It does not show cause and effect, seek to explain relationships, test hypothesis, make predictions, or get at meanings and implications.

DIFFUSION INDEX The percentage of a group of program indicators that are going up or down. This percentage then becomes the index. (46)

DIRECT COST (see also Indirect Cost) Are limited to those costs that are needed to immediately provide client services and that portion of overhead which varies directly with changes in provision of service. (e.g. payroll, monies for purchase of training, travel, etc.)

DISPLACEMENT Substitution of a rehabilitant for a non-handicapped worker. Jobs made available specifically to a Vocational Rehabilitation Agency are now unavailable to the general work force. (see also Social Cost)

ELEMENT Program classification subordinate to Program Level (*q.v.*) and Component (*q.v.*) (e.g. program management, staffing, *recruitment*).

EMPIRICAL DATA (see also Normative Data) Data obtained from observation, experiment, and past recorded experience without the use of theory or deduction (numerical counts of events that occurred, averages and ranges, etc.)

EXTERNALITIES Secondary effects, spill over effects, or spinoff, gained from contact with vocational rehabilitation, but not necessarily the achievement of vocational rehabilitation objectives. (e.g., achieving independence and responsibility, but not becoming self-supporting.)

EXPERIMENTAL STUDY (see also Observation, Description and Quasi Experimental) A study which investigates possible cause and effect relationships by exposing one or more groups to one or more treatment conditions and comparing the results to one or more control groups not receiving the treatment.

FIELD STUDY See Quasi Experimental Study.

HEURISTIC 1) Serving to guide, discover, or reveal. (115) 2) Solution of a problem by trial and error approach frequently involving the act of learning, and often leading to further discovery or conclusions without providing proof of the correctness of the outcome. (46)

HIGHER LEVEL EFFICIENCY CRITERIA Term used in the IRS manual which means measures of Effectiveness, Efficiency and Impact in this handbook.

HISTORICAL STUDY See retrospective.

HYPOTHESIS A statement of a presumed relationship between two variables one of which is considered the independent or "causal" variable which the other is viewed as the dependent or "effect" variable. (94)

INCIDENCE RATE (see also Prevalence Rate) Frequency of new occurrences of an event. (59) For instance, the Incidence Rate of welfare in a district's total population would be:

$$\frac{\text{New Welfare cases in 1972}}{\text{Total Population in 1972}}$$

INDEPENDENT VARIABLE (see also Control Variable, Dependent Variable and Intervening Variable) 1) Input, manipulated, or treatment variable; so called because they are "independent" of the outcome itself; instead, they are presumed to cause, effect or influence outcome. (48) 2) The "cause" in cause and effect relationships.

INDIRECT COSTS (see also Direct Cost) Are those costs that are not immediately needed to provide service. (e.g., overhead, administrative costs).

INTERVENING VARIABLE Variables that cannot be directly observed or measured and are hypothetical concepts intended to explain process between input and output. (48)

LEADING INDICATOR A time series of a program activity whose movement in a given direction precedes the movement of some other time series in the same direction (46) (e.g., few referrals in given time period leads to fewer status 26's at a later date).

MATRIX The framework of a table or chart.

MUTUALLY EXCLUSIVE CLASSES Classes so related that the membership in any one of them precludes the possibility of membership in any other class. (e.g., Male-Female; White-Nonwhite; ages 0 to 1, 2 to 3. . .)

NORMATIVE DATA (see also Empirical) 1) Represents the average performance or measure for a specified function of a specified homogeneous population. (e.g., high mechanical aptitude scores for carpenters, high IQ scores for physicians).

OBSERVATIONAL STUDY (see Descriptive and Experimental Studies) Personal observation of an activity which may use rating forms, check lists, or questionnaires to aid in data collection. (If analysis takes place this may be called Quasi Experimental).

ORGANIZATION An association, society, or administrative entity above program level (*q.v.*)

PANEL CONSENSUS The concluding agreement of a committee of experts concerning a given topic.

PREVALENCE RATE (see also Incidence Rate) Amount on hand, the number of people who have a given characteristic at a given moment. For example, the Prevalence Rate of AFDC would be expressed by:

$$\frac{\text{All AFDC in one city}}{\text{Total population of same city}} \quad (59)$$

PRIVATE COST (see also Social Cost) Total amount of money spent by an agency to provide services to the client, includes Direct Costs (*q.v.*) and Indirect Costs (*q.v.*) (used in Cost Benefit).

PROGRAM (see also Project and Organization) 1) Subordinate to organization, above project. 2) Part of the classification of Program Level (*q.v.*) (e.g. PA/VR, Section 2, etc.)

PROGRAM LEVEL A classification used in this manual used to distinguish between the categories of Program, Service Delivery, and Program Management.

PROJECT (see also Program and Organization) Subordinate to organization and program. (e.g. Vocational Rehabilitation Agency, Section 2 Program, *Alcoholism Project*).

PROSPECTIVE STUDY (see also Retrospective) 1) Looking forward in time. 2) Study proposed on events that have not yet occurred. (59) 3) Same as Experimental.

QUASI EXPERIMENTAL STUDY (see also Descriptive, Observational and Experimental) A study which attempts to approximate the conditions of an experimental study in a setting which does not allow the control and/or manipulation of all relevant variables. Most field experiments, operational research, and action research tend to fall into this category.

RAW DATA (see also Data) 1) Data that has not been processed. 2) The values or magnitude of observations before they have been subjected to any statistical treatment.

RETROSPECTIVE STUDY (see also Prospective) 1) Looking back in time. 2) A study using data that already exists. Can be used to suggest tenability of a hypothesis. 3) Data used based on past events, over which the evaluator has no control (e.g. before and after studies, interjurisdictional studies).

SET 1) A collection of elements having some feature in common, or which bear some relationship to one another. 2) A collection or family of objects considered as a whole.

SITE VISIT An official call on a program or project with the purpose of eliciting information.

SOCIAL COSTS (see also Private Cost, Displacement) The unanticipated and unseen costs due to disruption of society or some segment of society by the operation of a program or project (e.g. displacement of non-handicapped by the rehabilitants). Used in cost benefit analysis.

SOCIAL DISCOUNT RATE Is the rate at which future benefits are discounted given the project costs were invested elsewhere. (Discounting—the process of estimating the present value of perceived future benefits—the opposite of accruing interest). Thus, with a given discount rate of 5%, the projected earnings of \$1000 by a rehabilitant is presently worth \$950. Used in C/B studies.

VARIABLE (see also Control Variable, Dependent, Independent, Intervening Variable) 1) A trait quality or characteristics which exhibits differences in magnitude. 2) A quality which may occur in various magnitudes in different cases (e.g. age, years of employment, scores on standardized tests, amount of welfare benefits).

VISIONARY EVALUATION An assessment by a dreamer.

READING LIST

The following reading list has been compiled after extensive research in program evaluation. It is not presumed to be complete. It is presumed to be comprehensive and have utility as a guide for the program evaluator who is seeking a broader base of knowledge. To facilitate this process, the references have been catalogued under four headings: General Interest; Program Evaluation Design; Indices, Scales and Standard Nomenclature; and Bibliographies.

Contributions have been made by the Department of City and Regional Planning, University of California at Berkeley; the Social Welfare Research Institute at the University of Denver; and the Regional Rehabilitation Research Institute at the University of Oklahoma.

I. GENERAL INTEREST

Bennis, W.G., Theory and method in applying behavioral science to planned organizational change, *Operational Research and the Social Sciences*, ed. by J.R. Lawrence. London: Tavistock, 1966.

A really fine compact paper that covers the subject.

Benson, K.J. and D. Allen, Organizational structure and rehabilitation of the disadvantaged, *Rehabilitation and the Culturally Disadvantaged*, ed. by J.T. Kunce and C.S. Cope. Columbia, Missouri: University of Missouri Regional Rehabilitation Research Institute, 1969, pp. 172-201.

Organization theory applied to vocational rehabilitation.

Blau, P., Formal organizations—dimensions of analysis, *American Journal of Sociology* 63(7) (July, 1957) pp. 58-69.

Offers a conceptual framework for analysing formal organizations based on two theories; upward flow of demand (authority given up by subordinates), and statistical standards. The framework offers three areas for examination; group structure, organization structure and the developmental reactions (how change is handled).

Borgatta, E., Research problems in evaluation of health service demonstrations, *Milbank Memorial Quarterly* 44(2) (October, 1966) pp. 182-201.

The article is divided into seven sections that deal with some of the subtle pro and con issues in program evaluation. These include a utility theory for any program; the question of what is to be evaluated, that is, what the program is actually doing and what it is supposed to do? the problems of and rationalizations for negative findings; problems or rigors of experimental designs; advantages of looser designs, and the use of a program as an experimental manipulation. The author refers to program evaluation within the context of social research and deals with realistic problems of evaluation research.

Borus, M.E., and W.R. Tash, *Measuring the Impact of Manpower Programs; A Primer*. Ann Arbor, Michigan: University of Michigan and Wayne State University, Institute of Labor and Industrial Relations, 1970.

An excellent 76-page monograph discussing impact evaluations, methodologies, and indices. Stress is placed on cost/benefit analysis.

Buchanan, G.N. B.F. White and J.S. Wholey, *Political Considerations in the Design of Program Evaluation*. Paper presented at the American Sociological Association annual meeting, Denver, Colorado, September, 1971.

Buckley, W., Structural-functional analysis in modern sociology, *Modern Sociological Theory*, ed. by H. Becker and A. Boskoff. New York: Holt, Rhinehart and Winston, 1957.

Campbell, D.T., Considering the case against experimental evaluations of social innovations, *Administrative Science Quarterly* 15 (1) (March, 1970) pp. 110-112.

Arguments for the use of hard data.

Caro, F., *Readings in Evaluation Research*. New York: Russell Sage Foundation, 1971.

Thirty-one articles that will be of interest primarily to program evaluators and people interested in social organization research. Sections are devoted to issues in concepts of program evaluation, the difficulties of maintaining a viable evaluative department which contributes to decision-making within the organization and methodological case material. It is a good collection of representative articles.

Carter, G., The challenge of accountability—how we measure the outcome of our efforts, *Public Welfare* 29 (3) (Summer, 1971) pp. 267-277.

Good bibliography and good overview.

Churchman, C., *The Systems Approach*. New York: Delta Books, 1968.

Easy to read, explicit book on systems theory.

Deniston, O.L., I.M. Rosenstock, and V.A. Getting, Evaluation of program effectiveness, *Public Health Reports* 83 (4) (April, 1968) pp. 323-336. (Also in Schulberg, *Program Evaluation in the Health Fields*.)

This straightforward article suggests a method to evaluate program effectiveness. The author assumes that the program consists of a combination of resources, activities and objectives that interact. He presents a model that systematically describes resources, activities and objectives. Terms are defined. Then, the application of this model is shown with a discussion which tells how to describe the program by salesmanship and how and when to measure. The question is asked, "To what extent can achievement of the objective be attributed to the activities of the program?" Actual ratio formulae are suggested for analysis. A worthwhile overview article.

Deniston, O.L., I.M. Rosenstock, W. Welch, and V.A. Getting, Evaluation of program efficiency, *Public Health Reports* 83 (7) (July, 1968) pp. 603-610.

This is a companion article to the authors' previous article on program effectiveness. It covers a definition of effectiveness and efficiency (efficiency is a ratio between objectives obtained and actual resources used). The problem of quantity and quality, comparative uses of data for efficiency, the construction of efficiency curves, application of method and the limitations of measures of efficiency are discussed.

Donabedian, A., Evaluating the quality of medical care, *Milbank Memorial Quarterly* 44 (3) (pt. 2) (July, 1966) pp. 166-203. (Also in Schulberg's *Program Evaluation in the Health Fields*).

This paper is concerned with an evaluation of methodology. The case record

is cited as the principal source of information. Also noted are other directions for investigation, outcome, and the process and structure of the organization rendering the service. Specific topics discussed are sampling and selection, standards, measuring scales, reliability, bias, validity, and indices.

Donabedian, A., Frame of reference, *A Guide to Medical Care Administration*, Vol. II of *Medical Care Appraisal*. Washington, D.C.: American Public Health Association, 1969.

An administrators' discussion of a method for evaluating programs—the program's structure (inputs, resources, and organizational arrangements), the program's process (clinical techniques and services) and the program's outcome, i.e. well patients and satisfied patients.

Etzioni, A., *Modern Organizations*. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Etzioni, A., Two Approaches to organizational analysis: a critique and an analysis, *Administrative Science Quarterly* 5 (1960-1961) pp. 257-278. (Also in Schulberg's *Program Evaluation in the Health Fields*.)

The author presents two methodological approaches to program evaluation: the goal model or research approach, and the systems approach. He favors the systems approach because in theory it is more powerful and avoids value judgments. The author feels goal models select utopian criteria which are therefore open to value judgment. He defines two models in the systems approach: the survival evaluation and the effectiveness evaluation. The survival model shows that when the requirements are fulfilled the organization can exist. The effectiveness model presents several functional alternatives to the evaluation.

Fleck, A., Evaluation as a logical process, *Canadian Journal of Public Health* 52 (5) (May, 1961) pp. 185-191.

Dr. Fleck presents a vivid picture of the attitude directed toward implementing program evaluation in Civil Service. He describes the reason for resistance to evaluation as resistance to potential change. A section is set aside for the principles of program evaluation as used in the New York State Health Department: (1) Evaluation as a process has no end; (2) Programs are active in order to achieve three types of goals—community goals, administrative goals, and overt or covert goals; (3) Appraisal of benefits in the planning stage, and (4) Providing disease-control programs with an administrative basis.

Fleiss, J.L., *Statistical Methods for Rates and Proportions*. New York: Wiley, 1973.

Ghorpade, J., *Assessment of Organizational Effectiveness*. Pacific Palisades, California: Goodyear, 1971.

Greenberg, B.G., Evaluation of Social Programs, *Review of the International Statistical Institute* 36 (3) (1968) pp. 260-277.

(Also in Caro, *Readings in Evaluation Research*.)

A simulated clinical model of program evaluation.

Hage, J. and M. Aiken, *Social Change in Complex Organizations*. New York: Random House, 1970.

This excellent little paperback provides the steps and problems of changing complex organizations in easy-to-read terms. It is a must reader for people who seriously intend to measure change in program evaluation. Excellent bibliography.

Harbridge House, *Factors Influencing Agency Effectiveness: A Study for the Vocational Rehabilitation Administration*. Boston, Massachusetts: Harbridge House, 1963.

Hatry, H., E. Winnie and D.M. Fish, *Practical Program Evaluation for State and Local Government Officials*. Washington, D.C.: The Urban Institute, 1973.

A must for the program evaluator.

Hefferin, E.A. and E. Katz, Issues and orientations in the evaluation of rehabilitation programs; a review article, *Rehabilitation Literature* 32 (4) (April, 1971) pp. 98-113.

This paper has been prepared to present an overview of the major problems inherent in developing criteria for evaluating patient improvement, along with a review of empirical findings in various areas of rehabilitation research outcome.

Hills, W.G., *Evaluating Vocational Rehabilitation Programs*. Norman, Oklahoma: University of Oklahoma Regional Rehabilitation Research Institute, 1973.

Jahoda, M., and E. Barnitz, The nature of evaluation, *International Science Bulletin* 8 (3) (1955) pp. 353-363.

Johnson, R., F. Kast and J. Rosenweig, *The Theory and Management of Systems*. New York: McGraw-Hill, 1963.

The specific purpose of this text is to provide the administrator or manager with enough information to understand systems theory as it applies to his program management, and to help him analyze the operation as a system with manageable subsystems. It is easy to read and clear.

Kessner, D. and C. Kalk, *A Strategy for Evaluating Health Services*. Washington, D.C.: National Academy of Sciences, Institute of Medicine, 1973.

A methodological model presented for evaluating the quality of specific health services with six illustrations. The model, the tracer method, is readily adaptable to case review techniques in vocational rehabilitation.

Kunce, J.T. and C.S. Cope, *Rehabilitation and the Culturally Disadvantaged*. Columbia, Missouri: University of Missouri, Regional Rehabilitation Research Institute, Research Series No. 1, 1969.

While the topic of this short monograph was directed to the culturally disadvantaged, it is global in its applications. Of special interest are sections on counseling concerns and the disadvantaged program outcomes, and organizational implications. Must be read.

Lemkau, P.V. and B. Pasamanick, Problems in evaluation of mental health programs, *American Journal of Orthopsychiatry* 27 (1) (January, 1957) pp. 55-58.

Is the desire for global evaluation appropriate?

Levine, A.S., Cost benefit analysis of the work experience program—research strategy, *Welfare in Review* 4 (7) (August-September, 1966) pp. 1-9.

Levine, A.S., Evaluating program effectiveness and efficiency, *Welfare in Review* 5 (2) (February-March, 1967) pp. 1-7.

Levinson, Evaluation of social welfare programs, *Welfare in Review* 4 (10) (December, 1966), pp. 5-12.

An evaluator's discussion of the methodological approaches to evaluating programs—the research approach and the systems approach.

McMahon, B., T.H. Pugh, and G.B. Hutchinson, Principles in evaluation of community mental health programs, *American Journal of Public Health* 51 (7) (1971) pp. 963-968.

(Also in Schulberg's *Program Evaluation in the Health Fields*.)

This article defines two types of evaluative measures, those of accomplishment (effectiveness) and technique. It defines measurement variables for each and discusses the use of comparisons or controls.

Mangum, G.L., Perspectives on a positive manpower policy, *Social Economics for the 1970s*, ed. by G.F. Rohrlach. New York; Dunellen, 1970.

Mangum, G.L. and L.M. Glenn, *Vocational Rehabilitation and Federal Manpower Policy*. Washington, D.C.: National Manpower Policy Task Force and University of Michigan and Wayne State University, Institute of Labor and Industrial Relations, 1967.

Mark, J.A., Meanings and measures of productivity, *Public Administration Review* 32 (6) (November/December, 1972) pp. 747-753.

A discussion of how to measure productivity.

O'Toole, R. (ed), *The Organization, Management and Tactics of Social Research*. Cambridge, Massachusetts: Schenkman, 1971.

This book should be read by new evaluators. The collection is addressed to the knottier problems of how to run a social research group, gain access to information, do creditable research and survive, and then what to do with the results and how to get them to be used. It is very readable and practical. Gives examples to illustrate points.

Pelz, D. and F. Andrews, Detecting causal priorities in panel study data, *American Sociological Review* 28 (12) (December, 1964) pp. 836-848.

Rivlin, A.M., *Systematic Thinking for Social Action*. Washington, D.C.: The Brookings Institution, 1971.

Rossi, P.H., Practice, method and theory in evaluating social action programs, *Perspectives on Fighting Poverty*, ed. by J.L. Sundquist, New York: Basic Books, 1969.

Scanlon, J.W., *An Evaluation System to Support Planning Allocation and Control in a Decentralized, Comprehensive Manpower Program*. Washington, D.C.: The Urban Institute, 1971.

Schulberg, H. and F. Baker, Program evaluation models and the implementation of research findings, *American Journal of Public Health* 58 (9) (September, 1968) pp. 1248-1255.

Two models for studying program evaluation—goal attainment, or the research approach, and the systems approach are considered. The characteristics and limitations of each are dealt with as they affect implementation of research findings. In addition, the problems of implementing programmatic research findings are discussed and suggestions offered.

Schulberg, H., A. Sheldon and F. Baker, *Program Evaluation in the Health Fields*. New York: Behavioral Publications, 1969.

A collection of 35 articles that are of interest to program evaluators or anyone interested in the topic. Sections are devoted to general concepts and issues in program evaluation: various aspects of the quality and structure in methodology of program evaluation, various methodological techniques (PERT, Operations Research) and indices, examples of program evaluations and, finally, the problem of implementing the program evaluation findings. In comparing

Schulberg's and Caro's books, Schulberg draws heavily on the concrete thinking used in health evaluation. Caro deals with educational and social programs as examples in case materials section.

Sheps, M.C., Approaches to quality of hospital care, *Public Health Reports* 70 (9) (September, 1955) pp. 877-886.

This landmark article defines the purpose of the services, and the effects of the program on the quality of care. The article also defines the methods and consequent problems of evaluation as standards setting, evaluation of the elements of performance, estimating effects of care, and clinical evaluations of service (case review). Finally, there is a brief discussion of methods of comparisons.

Stein, H.D., M. Houghman and S.R. Zalba, Assessing social agency effectiveness: a goal model, *Welfare in Review* 6 (2) (1968) pp. 13-18.

Suchman, E.A., *Evaluative Research, Principles and Practices in Public Service and Social Action Programs*. New York: Russell Sage Foundation, 1967.

The author describes the techniques used to empirically determine the extent to which social goals are actually being achieved, to locate the barriers to achievement and to discover the unanticipated consequences of social actions. The book is divided into three main sections representing the conceptual, methodological, and administrative aspects of evaluation. Standard text.

Thompson, J.D., *Organizations in Action*. New York: McGraw Hill Book Co., 1967.

Excellent concise text of concepts in organizational management.

Trantow, D., An introduction to evaluation: program effectiveness and community needs, *Rehabilitation Literature*, 31 (1) (January, 1970) pp. 2-9.

This is an excellent article that integrates program evaluation with the program planning process. Must be read.

Tripodi, T., P. Fellin, and I. Epstein, *Social Program Evaluation Guidelines for Health, Education, and Welfare Administrators*. Itasca, Illinois: F. E. Peacock, 1971.

This 137-page book provides a framework for making decisions about program evaluation. It includes guidelines that can be used by administrators to initiate and plan evaluations and to effectively use evaluation consultants. Selected techniques are discussed in relation to their potential use in specific program evaluation. The book is intended for training administrators in health, education, and welfare agencies and as a text in courses in administration, planning educational social research (book jacket). It's an excellent overview but, read Hattry first.

U.S., DHEW, HSMHA; National Institute of Mental Health, *Planning for Creative Change in Mental Health, A Manual on Research Utilization*. DHEW, Publication number (HSM) 71-9059, 1973.

This 30-page document presents the ingredients and their relationships for the promotion of change. A model is presented to represent behavioral change. Twenty-one questions are given to answer the question of whether change is feasible. Terrific!

Weiss, C. H. The politicization of evaluation research, *Journal of Social Issues*, 26 (4) (1970) pp. 57-68.

A discussion of administrative repercussions following a program's evaluation.

Weiss, C. (ed), *Evaluating Action Programs*, Boston: Allyn and Bacon, 1972.

A truly superb reader, uniformly excellent, contains shorter articles by Suchman, Rossi, et al., Evans, etc., as well as many fine articles and papers not available elsewhere.

Wholey, J. S., et al., *Federal Evaluation Policy: Analyzing the Effect of Public Programs*. Washington, DC: The Urban Institute, 1970.

Assessment of evaluation activities of OEO, HEW, HUD, Labor.

II. PROGRAM EVALUATION DESIGN

Ackoff, R. and M.W. Sasieni, *Fundamentals of Operations Research*. New York: John Wiley, 1968.

Deals exclusively with problems involving linear, quadratic, and dynamic programming and PERT.

American Institute of Research, *Evaluative Research, Strategies and Methods*. Pittsburgh, Pennsylvania: American Institute of Research, 1970.

Anderson, S., Operations research in public health, *Public Health Reports* 79 (1964) pp. 297-305.

An outline and discussion for conducting a study using operational data.

Barsby, S., *Cost-Benefit Analysis and Manpower Programs*. Lexington, Massachusetts: Heath-Lexington Books, 1972.

Good basic text on cost-benefit analysis.

Berkowitz, M., PPBS, state-wide planning and the goals of vocational rehabilitation, *American Journal of Public Health* 58 (9) September, 1968) pp. 1633-1637.

Borus, M.E. and W.R. Tash, *Measuring the Impact of Manpower Programs; A Primer*. Ann Arbor, Michigan: University of Michigan and Wayne State University Institute of Labor and Industrial Relations, 1970.

An excellent 76-page monograph discussing impact evaluation, stressing cost-benefit analysis. List of criteria is offered.

Brown, P.L., *An Operational Model for a Planning-Programming-Budgeting System*. Paper presented at a Post Audit Seminar, Lexington, Kentucky, June, 1970.

This is a straightforward presentation of operationalizing PPBS.

Burton, T.L. and G.E. Cherry, *Social Research Technique for Planners*. London: G. Allen and Unwin, 1970.

Campbell, D.T., Factors relevant to the validity of experiments in social settings, *Psychological Bulletin* 12 (1957) pp. 297-312.

Campbell, D.T., Quasi-experimental design, *International Encyclopedia of Social Sciences*, Vol V. New York: MacMillan and Free Press, 1963.

A discussion of the prime research design used in program evaluation.

Campbell, D.T., Reforms as experiments, *American Psychologist* 24 (4) (April, 1969) pp. 409-429.

(Also in F. Caro, *Readings in Evaluation Research*.)

A discussion of nine different types of research designs.

Cochran, A.L. *Effectiveness, Efficiency*. Nuffield, Great Britain: Provincial Hospital Trust, 1971.

Cochran, W.C. *Sampling Techniques*. New York: John Wiley and Son, 1963.

Conley, R.W., Benefit-cost analysis and vocational rehabilitation, *Research Utilization in Rehabilitation Facilities*, ed. by R. Pacinelli. Washington, D.C.: International Association of Rehabilitation Facilities, 1971.

A fundamental discussion concerning four ways of doing cost-benefit and the measuring problems encountered.

Cornfield, J. and W. Haenszel, Some aspects of retrospective studies, *Journal of Chronic Diseases* 11 (5) (May, 1960) pp. 523-534.

This article defines the difference between causal and observational associations. It goes on to try to explain that it is not always possible to show casual association and so one must rely on observational associations. Next, a brief discussion on the measurement of risk occurrence (incidence, mortality, point

prevalence, and internal prevalence). Finally, he swings into a discussion on relation between prospective and retrospective studies, potential sources of error in retrospective studies, and measures of association. The examples given are epidemiological, but the method applies to any study using past events for data.

Crystal, R.A., and A.M. Brewster, Cost benefit and cost effectiveness analysis in the health field; an introduction, *Inquiry* 3 (4) (December, 1966) pp. 3-13.

A concise discussion of the techniques used in cost benefit and cost effectiveness analysis.

Davis, J.A., *Elementary Survey Analysis* Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Dawis, R.V., D.T. Hakes, G.W. England, and L.H. Lofquist, *Methodological Problems in Rehabilitation Research*. Minneapolis, Minnesota: University of Minnesota Industrial Relations Center, Minnesota Studies in Vocational Rehabilitation Bulletin No. 25, 1958.

Quick summary of methods in follow-up, population surveys, their reliability and validity.

Deeks, D.J. and A.J. Reynolds, Critical path method—a technique for project planning, *Management Controls* 13 (2) (Feb., 1966) pp. 24-32.

deNeufville, R., Systems analysis—a decision process, *Industrial Management Review* 11 (3) (Spring, 1970) pp. 49-58.

A discussion of systems analysis.

Donabedian, A., Evaluating the quality of medical care, *Milbank Memorial Quarterly* 44 (3) (July, 1966) (Part 2) pp. 166-203.

(Also in Schulberg, *Program Evaluation in the Health Fields*).

Discussion on data collection, selection, measuring standards, scales, bias, reliability and validity.

Donabedian, A., *A Guide to Medical Administration* Vol. II of *Medical Care Appraisal, Quality and Utilization*. New York: American Public Health Association, 1969.

This manual discusses methods of appraisals—a frame of reference, specific methods, issues of methods and technique, and issues of policy and implementation. Although the illustrations refer to medical care, the concepts are universal.

Eber, W.H., *Multivariate Analysis as a Rehabilitation Tool*. Washington, D.C.: Rehabilitation Services Administration, Joint Liaison Council of State Administrators, 1967.

Feldstein, M.S., The social time preference discount rate in cost benefit analysis, *Economic Journal* 74 (6) (June, 1964) pp. 360-379.

A clearly written article discussing techniques for establishing discount rates in cost benefit analysis.

Glidewell, J.C., I.N. Mensh, H.R. Domke, M. Gildea, and A.D. Buchmueller, Methods for community mental health research, *American Journal of Orthopsychiatry* 27 (1) (January, 1957) pp. 38-54.

Problems in hypothesis formation encountered in developing tests for effectiveness for one aspect of St. Louis County Health Department Mental Health Clinic for Children are presented. This article is based on a social research approach with control populations.

Gogsted, A., *Evaluation of Factors Determining the Results of Vocational Rehabilitation*. Baltimore, Maryland: Williams and Wilkins, 1968.

Step-by-step outline for follow-up studies.

Gyuro, S.J., The use of electronic data processing in an evaluation effort, *Evaluating MDTA Programs*. Lexington, Kentucky: University of Kentucky College of Education, Department of Vocational Education, 1971.

A summary of the use of computers applied to manpower programs.

Hatry, H., R.E. Winnie, and D. Fisk, *Practical Program Evaluation for State and Local Government Officials*. Washington, D.C.: The Urban Institute, 1973.

Excellent!

Herzog, A., and L. Denton, *A Model for the Evaluation of an Action Research Program*. Yarmouth, Nova Scotia: Nova Scotia NewStart Program, 1970.

This is a discussion model of how the author intends to evaluate programs. It is very abstract, but profitable for those who are willing to spend some time with it. The model is a quasi-research, systems-analysis model with effectiveness measures. Not for the beginner.

Herzog, E., *Some Guidelines for Evaluative Research*. Washington, D.C.: Department of Health, Education, and Welfare, Children's Bureau, 1959.

This report has been culled from the combined opinions expressed in 345 articles and studies concerned with the evaluation of effectiveness in psychotherapy. "It was written directly for the use of administrators who are considering setting up evaluative research in their agencies or are wanting to know how much reliance to put on the reported findings of such studies." The report includes chapters on the purposes of evaluation, which effects should be evaluated, methods used to assess change, what the findings mean, and practical implications. A goal-oriented social science research approach is outlined and critiqued.

Hinrichs, H.H. and G.M. Taylor, *Systematic Analysis: A Primer on Benefit-Cost Analysis and Program Evaluation*. Pacific Palisades, California: Goodyear, 1972.

Issac, S. and W.B. Michael, *Handbook in Research and Evaluation*. San Diego, California: Robert Knapp, 1971.

A good handbook with "how-to" information neatly summarized. Primary emphasis is on data analysis.

Institute on Rehabilitation Service. *Program Evaluation: A Beginning Statement*. Washington, D.C.: Department of Health, Education, and Welfare, Rehabilitation Services Administration, 1972.

Jacobs, D.O. *A Guide for Developing Questionnaire Items*. Alexandria, Virginia: Human Resources Research Organization, 1970.

This 21-page document's purpose is to present, in brief readable form, a guide to developing test questions for questionnaire instruments other than achievement tests. The guide was originally developed for the Army Infantry. Illustrative material is therefore defense oriented. The guide is based on Stanley Payne's book The Art of Asking Questions.

Kellman, H.R., Problems in measurement and evaluation of rehabilitations, *Archives of Physical Medicine and Rehabilitation* 43 (4) (April, 1962) pp. 172-181.

Kellman, H.R., Evaluation of rehabilitation for the long-term ill and disabled patient; some persistent research problems, *Journal of Chronic Disease and Disability* 17 (1964) pp. 631-639.

Kunce, Joseph T., The effectiveness of poverty programs: a review, *Rehabilitation and the Culturally Disadvantaged*, ed. by J.T. Kunce and C.S. Cope. Columbia, Missouri: University of Missouri Regional Rehabilitation Research Institute, 1969, pp. 127-170.

Kunce, J.T. Prediction and statistical overkill, *Measurement and Evaluation in Guidance* 4 (1) (April, 1971) pp. 38-42.

Lyden. F.G. and E.G. Miller, *Planning, Programming, Budgeting*. Chicago, Illinois: Markham, 1968.

Mann, J., Technical and social difficulties in conduct of evaluative research, *Changing Human Behavior*. New York: Charles Scribner's Sons, 1965.
(Also in F. Caro, *Readings in Evaluative Research*.)

An excellent discussion of the technical and administrative problems encountered in conducting research in an institutional setting.

McMahon, B., T.F. Pugh, and G.B. Hutchinson, Principles in the evaluation of community mental health programs, *American Journal of Public Health* 51 (7) (1961) pp. 963-968.

A clinician's discussion of methods for evaluating programs—accomplishment (effectiveness) and technique (process).

McMahon, B., T.F. Pugh, and J. Ispen, *Epidemiologic Methods* Boston: Little, Brown and Co., 1960.

Standard text used in defining epidemiological methods. Easily applicable method for surveys and studies that use measures other than behavioral measures.

Mayer, R. and P. Pipe, *Analyzing Performance Problems*. Belmont, California: Fearon, 1970.

McGuigan, E.J., *Experimental Psychology: A Methodological Approach*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

This easy-to-read text defines the steps in experimental designs, the major elements of an experiment, and four possible research designs used in experimental psychology. Students who have used this text testify to its utility.

McKinley, J.B., *Research Methods in Health Care*. New York: Prodist, 1974.

A collection of articles reprinted from the Milbank Memorial Quarterly concerning issues surrounding research methodology. The majority of these articles are addressed to problems in survey technique.

Miller, D., *Handbook of Research Design and Social Measurement*. New York: David McKay, 1964.

This handbook is not a text nor is it a substitute for one. It is a compilation of resources organized to provide reference to the essential materials needed in the design of research. The three sections include information on the steps for doing research: the mathematical tests that may be used and a representative selection of forty sociometric scales and indexes with a description of each. Excellent! Everybody should have one.

Morehouse, T., *The Problem of Measuring the Impact of Social Action Programs*. College, Alaska: University of Alaska Institute of Social, Economic, and Government Research, 1972.

Nay, J.N., *Benefits and Costs of Manpower Training Programs: A Synthesis of Previous Studies with Reservations and Recommendations*. Washington, D.C.: The Urban Institute, 1971.

Oppenheim, A.N., *Questionnaire Design and Attitude Measurement*. New York: Basic Books, 1966.

This short book is addressed to students of marketing and advertising research staffs, and to any professional who needs to evaluate or plan social research. It outlines in step-by-step form the procedures for the design of questionnaires and surveys and techniques for the measurement of attitudes. In addition, each chapter has annotated bibliographic citations for additional information.

O'Toole, R., *The Organization, Management, and Tactics of Social Research*. Cambridge, Massachusetts: Schenkman Publishing Co., 1971.

Richardson, F., Peer review of medical care, *Medical Care* 10 (1) (1972) pp. 29-39.

Roman, D.D., The PERT system: an appraisal of program evaluation review technique, *Journal of Academy of Management* 5 (1962) pp. 57-65.

(Also in H. Schulberg, *Program Evaluation in the Health Fields*.)

An excellent summary of what PERT is and can do.

Rosenstock, I., Some principles of research design in public health, *American Journal of Public Health* 51 (2) (February, 1961) pp. 266-277.

This paper presents a very practical set of principles for research design that can guide the planning of investigations. They are illustrated in terms of a project actually carried out by the author.

Siegel, S., *Non Parametric Statistics*. New York: McGraw-Hill, 1956.

This is one of the standard texts in nonparametric statistics.

Somers, G.C. and W.D. Wood, *Cost-Benefit Analysis of Manpower Programs; Proceedings of a North American Conference*. Madison, Wisconsin: University of Wisconsin Center for Studies in Vocational and Technical Education, and Queen's University, Industrial Relations Centre, 1968.

Stromsforfer, E.W. *Review and Synthesis of Cost-Effectiveness Studies of Vocational and Technical Education*. Columbus, Ohio: Ohio State University, ERIC Clearinghouse on Vocational and Technical Education, 1972.

*This is a first-rate overview of cost-benefit and cost-effectiveness studies. It should be read by all social scientists who are interested in cost-effectiveness studies. Note must be taken, however, that certain precisely defined economic terms are used and not explained. These simple concepts may be secured from any fundamental micro-economics textbook (such as Mansfield, E., *Micro Economics*, New York: Norton, 1970). Within this monograph is a 36-page categorized bibliography, which includes a 29-item list of references to other cost-benefit/cost-effectiveness bibliographies.*

Suchman, E., *Evaluative Research, Principles and Practices in Public Service and Social Action Programs*. New York: Russell Sage Foundation, 1967.

Suchman, E. A model for research and evaluation in rehabilitation, *Sociology and Rehabilitation*, ed. by M.B. Sussman. Washington, D.C.: American Sociological Association, 1965.

Suchman describes two models for evaluation of rehabilitation. The first is a basic research approach using rehabilitation as an intervening variable which modifies the relationship between disability (cause) and the consequences of disability as the dependent variable (client). The second model is an applied evaluative model for the determination of effectiveness in rehabilitation agencies, that is, the degree to which the activities and the staff used are effective in changing conditions and behavior.

Trantnow, D.J., An introduction to evaluation, program effectiveness, and community need, *Rehabilitation Literature* 31 (1) (January, 1970) pp. 2-9.

Warmbrod, R., Interpreting evaluation findings, *Evaluating MDTA Programs*, ed. by F.L. McKinney and A.J. Mannebach. Lexington, Kentucky: University of Kentucky, Research Coordinating Unit, 1971, pp. 279-295.

A discussion of rules for interpreting findings.

Weiss, C.H. *Evaluation Research: Methods of Assessing Program Effectiveness*. Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

III. INDICES, SCALES, AND STANDARD NOMENCLATURE

American Medical Association, Committee on Ratings of Mental and Physical Impairment, *Guides to the Evaluation of Permanent Impairment*. Chicago: 1971.

Barry, J.R. and M.R. Malinovsky, *Client Motivation for Rehabilitation; A Review*. Gainesville, Florida: University of Florida Regional Rehabilitation Research Institute, 1968.

Bateman, W., Assessing program effectiveness: a rating system for identifying relative project success, *Welfare in Review* 6 (1) (1968) pp. 1-10.

A proposed set of variables that could be used for national comparisons.

Berkowitz, M., *Weighted Closures in Vocational Rehabilitation*, unpublished paper, New Brunswick, New Jersey: Rutgers University, Bureau of Economic Research, no date.

Bledsoe, R.C., D.R. Denny, C.D. Hobbs, and R.S. Long, Productivity management in the California social service program, *Public Administration Review* 32 (6) (November-December, 1972) pp. 799-803.

This article relates the experience of the California Community Service's "contract" service provision to welfare clients. Through this approach both the client and the worker establish a goal and time frame.

Cantril, H., Job satisfaction and family functioning, *Patterns of Human Concerns*. New Brunswick, New Jersey: Rutgers University Press, 1965.

Ciocco, A., On indices for the appraisal of health department activities, *Journal of Chronic Diseases* 11 (1961) pp. 509-522.

(Also in H. C. Schulberg, *Program Evaluation in the Health Fields*.)

This article offers a brief history and discussion of standard measures and comparisons in health that have withstood the test of time.

Institute for Rehabilitation Services, *Training Guide in Caseload Management for Vocational Rehabilitation Staff*. Washington, D.C.: Department of Health, Education, and Welfare, Vocational Rehabilitation Administration, 1965.

This manual considers all variables and their interrelationships relevant to evaluating counselor performance.

Kessner, D.M. and C.E. Kalk, Development of the tracer method, *A Strategy for Evaluating Health Services*, Washington, D.C.: National Academy of Sciences, Institute of Medicine, 1973.

Proposed technique for identifying those diseases whose course of treatment is sufficiently outlined, and, therefore, can serve as criteria in evaluating program process.

Kirsecuk, T., *Goal Attainment Scaling*. Minneapolis, Minnesota: Program Evaluation Project, 1969.

A system used to describe and evaluate patient problems in mental health centers. This measurement of client behavior change looks promising as a way of comparing stated client goals with client success.

Likert, R., Profile of organizational characteristics, *The Human Organization: Its Management and Value*. New York: McGraw-Hill, 1967.

Marvin, P., *Management Goals: Guidelines and Accountability*. Homewood, Illinois: Dow, Jones, and Irwin, 1968.

Miller, D.C., *Handbook of Research Design and Social Measurement*. New York: David McKay, 1967.

Included in this handbook are forty descriptions of scales and indices. The topics cover social status, group structure, morale and job satisfaction, community and social participation, leadership in organizations, important attitude scales, family and marriage, personality measurement, and, most importantly, an inventory of measures utilized in the American Sociological Review, 1951-1960, with original citations.

Miller, L.A. and M. Barillas, Using weighted 26 closures as a more adequate measure of counselor and agency effort, *Rehabilitation Counseling Bulletin* 11 (12) (December, 1967) pp. 117-121.

Miller, L. A. and R. Roberts, *Rehabilitation Counselor Change Associated with Experimental Continuing Education*. Iowa City: University of Iowa College of Education, 1971.

Findings relevant to inservice training.

Minnesota, University of, Industrial Relations Center, *Manual for the Minnesota Employment Satisfactoriness Scale*, by D. L. Gibson. Minneapolis, Minnesota: Minnesota Studies in Vocational Rehabilitation, Volume 27, 1970.

Minnesota, University of, Industrial Relations Center, *Manual for the Minnesota Satisfaction Scales*. Minneapolis, Minnesota: Minnesota Studies in Vocational Rehabilitation, Bulletin 45, 1967.

Muthard, J.E. and L. Miller, *The Criteria Problem in Rehabilitation Counseling*. Iowa City: University of Iowa, College of Education, 1966.

The thrust of this monograph is directed to establishing valid, reliable criteria to evaluate counselor performance. Some are based on actual performance criteria, while others are based on opinion scales.

Muthard, J.E. and L. Miller, *The Rehabilitation Counselor Rating Scale*. Washington, D.C.: Joint Liason Committee of the Council of State Administrators of Vocational Rehabilitation, Studies in Rehabilitation Counselor Training No. 6, 1968.

Nagi, S.Z. *Disability and Rehabilitation: Legal, Clinical, and Self-Concepts and Measurement*. Columbus, Ohio: Ohio State University Press, 1969.

Noble, J.H., Actuarial system for weighting case closures, *Rehabilitation Record* 14 (5) (September-October, 1973) pp. 34-37.

Oklahoma, University of, Regional Rehabilitation Research Institute, *Management Practices in Vocational Rehabilitation District Offices*, by H. Vialle and W.H. Hills, Norman. Oklahoma: 1973.

The only study of vocational rehabilitation management practice.

Oklahoma, University of, Regional Rehabilitation Research Institute, *Third Party Agreements; Program Evaluation*. Norman, Oklahoma: 1973.

Reagles, K. *Human Service Systems I*. Madison, Wisconsin: Human Service Systems, 1973.

Measures the degree of change experienced by clients served through various human services.

Richardson, W., *Rehabilitation Counselor Sub Role Adaptation Scale*. Fayetteville, Arkansas, University of Arkansas Research and Training Center, 1973.

A method to evaluate counselor behavior during client interview.

Sermon, D., *The Difficulty Index: An Expanded Measure of Counselor Performance*. St. Paul, Minnesota: Minnesota Division of Vocational Rehabilitation, 1972.

Texas Rehabilitation Commission. *Handbook for Program Evaluation*. Austin, Texas: 1972.

This draft for Texas' program evaluation outlines a step-by-step procedure for the evaluator. The areas specifically examined are case management (case distribution, case flow, time in status) and case record review (pre and post). In addition, other areas for evaluation are suggested, such as administrative functions of district offices and analysis of closure status.

Thompson, E.T. and A.C. Hayden, *Standard Nomenclature of Disease and Operations*. New York: McGraw-Hill, 1961.

U.S. Department of Commerce. Bureau of Census, *Statistical Abstract of the United States*. Washington, D.C.: United States Government Printing Office, annual.

U.S. Department of Health, Education, and Welfare, Public Health Service, National Center for Health Statistics, *Current Listing and Topical Index to the Vital and Health Statistics 1962-1972*. Washington, D.C.: United States Government Printing Office, May, 1973.

This can be ordered from the Office of Information, National Center for Health Statistics, Public Health Service, Rockville, Maryland 20852 (No. (HSM) 73-1301 revised). Of particular interest to vocational rehabilitation is Series 10.

U.S. Department of Health, Education, and Welfare. Public Health Service, National Center for Health Statistics, Definitions of certain terms used in this

report, in *Current Estimates from the Health Interviews Survey*. Washington, D.C.: United States Government Printing Office, 1972.

A concise list of disability definitions used in Census material.

U.S. Department of Health, Education, and Welfare, Social Security Administration, Office of Research and Statistics, *Survey of Disabled Adults*, Washington, D.C.: United States Government Printing Office, 1966.

Various report topics.

U.S. Department of Health, Education, and Welfare, Social and Rehabilitation Service, Rehabilitation Services Administration, *Caseload Statistics; State Vocational Rehabilitation Agencies*. Washington, D.C.: United States Government Printing Office, annual.

U.S. Department of Labor, Manpower Administration, *Dictionary of Occupational Titles, Vol II. Occupational Classifications*. Washington, D.C.: United States Government Printing Office, 1965.

U.S. Department of Labor, Bureau of Labor Statistics, *NEWS*. Washington, D.C. Department of Labor, Office of Information, monthly.

Publications on the employment situation, job vacancies, and productivity and costs in the private economy.

U.S. Executive Office, Bureau of the Budget, *Measuring Productivity of Federal Government Organizations*. Washington, D.C.: United States Government Printing Office, 1964.

Westerheide, W.J. and L. Lenhart, *Case Difficulty and Client Change: A Review of the Literature*. Oklahoma City, Oklahoma: Oklahoma Department of Institutions, Social, and Rehabilitation Services, 1974.

Wheeler, D., *The Measure of Job Relatedness for Vocational Evaluation*. St. Paul, Minnesota: Minnesota Research Coordinating Unit for Occupational Education, 1971.

Wisconsin, University of, Regional Rehabilitation Research Institute, *Wisconsin Studies in Vocational Rehabilitation Monographs II-X*. Madison, Wisconsin: 1968.

A summary and collection of studies surrounding the issue of counselor perceptions and attitudes as they affect performance.

Wisconsin, University of, Regional Rehabilitation Research Institute, *A Scale of Rehabilitation Gain for Clients of an Expanded Vocational Rehabilitation Program*, by K. Reagles. Madison, Wisconsin: Monograph XIII in Wisconsin Studies in Vocational Rehabilitation, 1970.

IV. BIBLIOGRAPHIES

American Medical Association, *Utilization Review: A Handbook for Medical Staff*. Chicago, Illinois: 1968.

Anderson, A.J. Methodology in evaluating the quality of medical care: a selected annotated bibliography 1962-1968, *A Guide to Medical Care Administration, Vol II. Medical Care Appraisal—Quality and Utilization*, by A. Donnebedian. Washington, D.C.: American Public Health Association, 1969.

Anderson, A.J. and I. Altman, *Methodology in Evaluating the Quality of Medical Care: An Annotated Selected Bibliography 1955-1961*. Pittsburgh, Pennsylvania: University of Pittsburgh Press, 1962.

Dent, J.K., *A Bibliographic Index of Evaluation in Mental Health*. Washington, D.C.: United States Government Printing Office, Public Health Service Publication No. 1545, 1966.

DeGeyndt, W. and K.B. Ross, *Evaluation of Health Programs—An Annotated Bibliography*. Minneapolis, Minnesota: Systems Development Project, no date.

Eichhorn, R., *The Utilization of Health Services Indices and Correlates—A Research Bibliography*. Lafayette, Indiana: Purdue University, 1972.

Gupta, V., *Evaluation of Technical Quality of Research—An Annotated Bibliography*. Minneapolis, Minnesota: Systems Research, Inc., 1973.

Hills, W.G., *Evaluating Vocational Rehabilitation Programs*. Norman, Oklahoma: University of Oklahoma Regional Rehabilitation Research Institute, 1973.

Program Evaluation Project, *A Bibliographic Index to Evaluation in Mental Health*. Minneapolis, Minnesota: 1972.

Stromsdorfer, E.W., *Review and Synthesis of Cost-Effectiveness Studies of Vocational and Technical Education*. Columbus, Ohio: ERIC Clearinghouse on Vocational and Technical Education, 1972.

U.S. Department of Health, Education, and Welfare, Public Health Service, National Center for Health Statistics, *Annotated Bibliography on Vital and Health Statistics*. Washington, D.C.: United States Government Printing Office, 1970.

U.S. Department of Health, Education, and Welfare, Public Health Service,

National Center for Health Statistics, *Catalogue of Publications of the National Center for Health Statistics*. Washington, D.C.: United States Government Printing Office, HSM 73-1306, 1972.

Weitzner, M., L.E. Bellin, F. Kavalier, and L. Feldman, *Cost Control and Utilization Review in Health Settings: A Selected Annotated Bibliography*. New York: 1972.

REFERENCE LIST

1. Ackoff, R.L. and M. W. Sasieni, *Operations Research*. New York: John Wiley, 1968.
2. American Public Health Association, *A Guide to Medical Care, Vol I*. Washington, D.C.: 1965.
3. Appel, G.L. and P.C. Schenker, An analysis of Michigan's experience with work incentives, *Monthly Labor Review* 94 (9) (September, 1971).
4. Auerbach Corporation, *Part I: Technical Proposal to Conduct an Evaluation of the Welfare Demonstration Project of the Public Employment Program*. Philadelphia, Pennsylvania: 1971.
5. Auerbach Corporation, *An Impact Evaluation of the Work Incentive Program (WIN) Vol I. Major Findings and Conclusions*. Philadelphia, Pennsylvania: DL 50 40 69 02, 1972.
6. Barsby, S., *Cost Benefit Analysis and Manpower Programs*. Lexington, Massachusetts: Heath-Lexington, 1972.
7. Baumheier, E.C., *A Conceptual Framework for the Description and Analysis of Human Service Delivery Systems*. Denver, Colorado: University of Denver, Social Welfare Research Institute, 1972.
8. Bax, J.A., Social services: today and tomorrow, *Public Welfare* 30 (1) (Winter, 1972).
9. Bercaw, W.W., *A Systems Analysis of the MDTA Institutional Training Program*. Anaheim, California: North American Information Systems Company, 1971.
10. Borus, M.E., A benefit-cost analysis of the economic effectiveness of re-training the unemployed, *Yale Economic Essays*. New Haven, Connecticut: Yale University Press, 1964.
11. Borus, M.E. and W.R. Tash, *Measuring the Impact of Social Action Programs: A Primer*. Ann Arbor, Michigan: University of Michigan and Wayne State University, Institute for Labor and Industrial Relations, 1970.
12. Buckley, W., Structural functional analysis in modern sociology, *Modern*

Sociological Theory, ed. by N. Becker and A. Boskoff. New York: Rhinehart and Winston, 1957.

13. California, Department of Rehabilitation, *Placement Task Force Report*, by G. Allen, J. Devine, J. Erickson, B. Hancock, T. Hickey, J. Morales, R. Noble and P. Seldon. Sacramento, California: 1971.
14. California, Department of Rehabilitation, *An Individualized Approach to the Problems of Poverty and Dependency*, by Z.L. Gulledge. Sacramento, California: RD 2089, 1970.
15. California, Department of Rehabilitation, *Meanwhile, At Home . . . ; A Follow-Up Study of Mentally Ill Vocational Rehabilitation Clients*, by F. Schlamp. Sacramento, California: RD 1555, 1971.
16. California, Department of Rehabilitation, *The Educational Level of Rehabilitation Counselors as Related to Performance*, by P.F. Mueller and I. Atlas. Sacramento, California: 1970.
17. California, Department of Rehabilitation, *Rehabilitation Program Review Report: Statewide Report on Findings from District Reviews*, by B. Dieckman. Sacramento, California: 1970.
18. Cantril, H., *The Pattern of Human Concerns*. New Brunswick, New Jersey: Rutgers University Press, 1965.
19. Churchman, C.W., *The Systems Approach*. New York: Delta Books, 1968.
20. Ciocco, A., On indices for the appraisal of health department activities, *Journal of Chronic Diseases* 11 (May, 1960).
21. Collignon, F., A. Zawada, B. Thompson, and J. Markowitz, *Guidelines and Criteria for Evaluating Vocational Rehabilitation, Tenth Institute on Rehabilitation Services*. Berkeley, California: University of California at Berkeley, Institute of Urban and Regional Development, 1972.
22. Colorado, Division of Vocational Rehabilitation, *Client Case Control and Update Months in Current Status*. Denver, Colorado: 1972.
23. Conley, R.W., A benefit-cost analysis of the vocational rehabilitation program, *Journal of Human Resources*, 4 (2) (Spring, 1969).
24. Connecticut, Division of Vocational Rehabilitation, *Status Count*. Hartford, Connecticut: 25-P-65443, 1970.

25. DeNeufville, R., Systems analysis—a decision process, *Industrial Managers' Review* 11 (3) (Spring, 1970).
26. Deniston, O.L., I.M. Rosenstock, W. Welch, and V.A. Getting, Evaluation of program efficiency, *Public Health Reports* 83 (7) (July, 1968).
27. Deniston, O.L., I.M. Rosenstock and V.A. Getting, Evaluation of program effectiveness, *Public Health Reports* 83 (4) (April, 1968).
28. Derr, J.M. and E.C. Baumheier, *Systems Approach to the Evaluation of Social Intervention Programs*. Denver, Colorado: University of Denver Social Welfare Research Institute, May, 1973.
29. Donabedian, A., Evaluating the quality of medical care, *Milbank Memorial Quarterly* 44 (3) (Part 2) (July, 1966).
30. Donabedian, A., *A Guide to Medical Care Administration Vol. II. Medical Care Appraisal—Quality and Utilization*. New York: American Public Health Association, 1969.
31. Ettinger, M.L., The employability of the epileptic, *Rehabilitation Research*, ed. by G. Wright. Madison, Wisconsin: University of Wisconsin, 1968.
32. Etzioni, A., Two approaches to organizational analysis: a critique and a suggestion, *Administrative Science Quarterly* 5 (May, 1960).
33. Fleck, A.C., Evaluation as a logical process, *Canadian Journal of Public Health* 52 (5) (1961).
34. Florida, Division of Vocational Rehabilitation, *Florida Difficulty Index in Achieving Rehabilitation*. Tallahassee, Florida: 1973.
35. Florida, Division of Vocational Rehabilitation, *Counselor Study—A Study of the Time Spent by Florida Counselors in Performing their Duties and of the Relationship Between High Achievement and Quality of Work*, by H.A. Meyer. Tallahassee, Florida: 1971.
36. Florida, Division of Vocational Rehabilitation, *Caseload Analysis: High and Low Achievers*, Florida: 1971.
37. Freund, J.E. and F.J. Williams, *Elementary Business Statistics: The Modern Approach*. Englewood Cliffs, New Jersey: Prentice-Hall, 1964.

38. Goldin, G., Rehabilitation Counseling criteria, *Rehabilitation Record* 5 (5) (September-October, 1964).
39. Grigg, C., *Vocational Rehabilitation of Disabled Public Assistance Clients*. Tallahassee, Florida: Florida State University, Institute for Social Research, 1969.
40. Grigg, C.M., A.B. Holtzmann, and P.Y. Martin, *Vocational Rehabilitation of Disabled Public Assistance Clients: An Evaluation of Fourteen Research and Demonstration Projects*. Tallahassee, Florida: Florida State University, Institute of Social Research, 1969.
41. Hatry, H., R.E. Winnie, and D. Fisk, *Practical Program Evaluation for State and Local Government*. Washington, D.C.: The Urban Institute, 1973.
42. Hays, W., *Statistics for Psychologists*. New York: Holt, Rhinehart, and Winston, 1965.
43. Heckert, J.B. and J.B. Wilson, *Business Budgeting and Control*. New York: Ronald Press, 1967.
44. Herzog, A. and L. Denton, *A Model for the Evaluation of an Action Research Program*. Yarmouth, Nova Scotia: Nova Scotia NewStart Program, 1970.
45. Illinois, Institute for Social Policy, Welfare Division, *Evaluation Design and Methodology*. Chicago, Illinois: University of Illinois, 1972.
46. Institute on Rehabilitation Services, *Program Evaluation—A Beginning Statement*. Washington, D.C.: Department of Health, Education, and Welfare, Social and Rehabilitation Services, Rehabilitation Services Administration, 1972.
47. Institute on Rehabilitation Services, *Training Guide in Caseload Management for Vocational Rehabilitation Staff*. Washington, D.C.: Department of Health, Education, and Welfare, Vocational Rehabilitation Administration, 1965.
48. Issac, S. and W.B. Michael, *Handbook in Research and Evaluation*. San Diego, California: Robert Knapp, 1971.
49. James, G., Planning and evaluation of health programs, *Administration of*

Community Health Services, ed. by E. Confrey. Chicago, Illinois: International City Managers' Association, 1961.

50. Johnson, R.A., F.E. Kast, and J.E. Rosenweig, *The Theory and Management of Systems*. New York: McGraw-Hill, 1963.
51. Johnson, R.W., Number of interviews, diagnosis and success of counseling, *Journal of Counseling Psychology* 12 (3) (1965).
52. Katz, A. and Levin, Traditions in research of the diffusion of innovation, *American Sociological Review* 28 (2) (April, 1963).
53. Likert, R., *The Human Organization: Management and Value*. New York: McGraw-Hill, 1963.
54. Kiresuk, T.J. and R.E. Sherman, *Goal Attainment Scaling: A General Method for Evaluating Comprehensive Mental Health Programs*. Minneapolis, Minnesota: Program Evaluation Project, 1972.
55. Kramer, L., *An Evaluation of Youth Programs*. Washington, D.C.: National Conference of Christians and Jews, 1972.
56. Kunce, J. T., R. Mahoney, R. Campbell, and J. Findlay, *Rehabilitation in the Concrete Jungle*. Columbia, Missouri: University of Missouri, Regional Rehabilitation Research Institute, 1969.
57. Kunce, J.T. and C.S. Cope, *Rehabilitation and the Culturally Disadvantaged*. Columbia, Missouri: University of Missouri, Regional Rehabilitation Research Institute, 1969.
58. Kutner, B. et al., *Five Hundred Over Sixty*. New York: Russell Sage Foundation, 1956.
59. MacMahon, B., T.F. Pugh, and J. Ipsen, *Epidemiologic Methods*. Boston, Massachusetts: Little, Brown, and Co., 1960.
60. MacMahon, B., T.F. Pugh, and G.B. Hutchinson, Principles in evaluating community mental health programs, *American Journal of Public Health* 51 (1961).
61. McGuignon, E.J., *Experimental Psychology: A Methodological Approach*. Englewood Cliffs, New Jersey, Prentice-Hall, 1968.

62. Marvin, P., *Management Goals: Guidelines and Accountability*. Homewood, Illinois: Dow Jones-Irwin, 1968.
63. Massachusetts, Rehabilitation Commission, Research Unit, *An Evaluation of an Expansion Project to Rehabilitate Public Assistance Recipients Using Vocational Rehabilitation Counselor-Social Worker Teams*, by A. Lipson. Boston, Massachusetts: 1973.
64. Michigan, Department of Social Services, *Demonstration Project Coordinating Welfare and Rehabilitation for the Disabled Public Assistance Recipient*, by B. Houston. Lansing, Michigan: 1972.
65. Miller, D.C., *Handbook of Research Design and Social Measurement*. New York: David McKay, 1967.
66. Miller, L.A. and M. Barillas, Using weighted 26 closures as a more adequate measure of counselor and agency effort, *Rehabilitation Counseling Bulletin* 11 (2) (1967).
67. Minnesota, Division of Vocational Rehabilitation, *Monthly Report*. St. Paul: 1972.
68. Morehouse, T., *The Problem of Measuring the Impact of Social Action Programs*. College, Alaska: University of Alaska, Institute of Social, Economic, and Government Research, 1972.
69. Murphy, J., et al., *An Evaluation of the Messiah-Willard Day Care Center 1115 Demonstration Project*. Minneapolis, Minnesota: Institute for Interdisciplinary Studies, 1972.
70. Muthard, J.E., *Monitoring, Evaluation, and Research*. Gainesville, Florida: University of Florida, Regional Rehabilitation Research Institute, 1971.
71. Muthard, J.E. and L.A. Miller, *The Rehabilitation Counselor Rating Scale*. Washington, D.C.: Joint Liason Committee of the Council of State Administrators of Vocational Rehabilitation, Studies in Rehabilitation Counselor Training No. 6., 1968.
72. Muthard, J.E. and L.A. Miller, The structured case review study, *The Criteria Problem in Rehabilitation Counseling*. Iowa City: University of Iowa, College of Education, 1966.
73. Nevada, Rehabilitation Division, *Expansion Grant. Goals. (Appendix A)*

Training Needs Assessment for Counseling Staff. Carson City, Nevada: 25-P-65408/9-01, 1971.

74. New Jersey, *Weighted Closures*, by M. Berkowitz. New Brunswick, New Jersey: Rutgers University, Bureau of Economic Research, 1972.
75. Noble, J.H., Actuarial system for weighting case closures, *Rehabilitation Record* 14 (5) (September-October, 1973).
76. North Carolina Division of Vocational Rehabilitation, *Program of Evaluation*. Raleigh, North Carolina: mimeo, 1971.
77. Ohio State University, *Plan for Evaluating the PA/VR Franklin County Expansion Grant Project*, by L. Schneiderman. Columbus, Ohio: 1972.
78. Oklahoma, Department of Institutions, Social and Rehabilitation Services, *State-of-the-Art on Weighted Closures*. Oklahoma City, Oklahoma: 1971.
79. Oppenheim, A.N., *Questionnaire Design and Attitude Measurement*. New York: Basic Books, 1966.
80. Oregon, Vocational Rehabilitation Division, *A Program Information and Evaluation Systems Model for Rehabilitation Agencies*, by D. Schausten. Salem, Oregon: mimeo, 1971.
81. Pacific Training and Technical Assistance Corporation, *The Impact of Urban WIN Programs*. Washington, D.C.: DL 51 09 70 10, May, 1972.
82. Raichur, L. and E.C. Baumheier, *Overall Comments on the Cost-Benefit/Cost-Effectiveness Approach*. Denver, Colorado: University of Denver Social Welfare Research Institute, May, 1973.
83. Rivlin, A.M., *Systematic Thinking for Social Action*. Washington, D.C.: The Brookings Institution, 1971.
84. Rosenstock, I.M., Some principles of research design in public health, *American Journal of Public Health* 51 (1961).
85. Sankovsky, R. Some principles of research, *Rehabilitation Research and Practice Review* 3 (2) (Special Issue), 1972.
86. Schaefer, I.J., A follow-up of 73 psychiatric rehabilitation clients, *Rehabilitation Research and Practice Review* 3 (2) (Special Issue) 1972.

87. Schulberg, H. and F. Baker, Program evaluation models and the implementation of research findings, *American Journal of Public Health* 58 (9) (1968).
88. Sermon, D., *The Difficulty Index: An Expanded Measure of Counselor Performance*. St. Paul, Minnesota: Minnesota Division of Vocational Rehabilitation, 1972.
89. Sewell, D.O., *Training the Poor: A Cost Benefit Analysis of Manpower Programs in the U.S. Antipoverty Program*, Kingston, Ontario: Queen's University, Industrial Relations Centre, 1971.
90. Sheps, M.C., Approaches to the quality of hospital care, *Public Health Reports* 70 (9) (1955).
91. Siegel, S., *Non Parametric Statistics*. New York: McGraw-Hill, 1956.
92. Spencer, G., *Structure and Dynamics of Social Intervention*. Lexington, Massachusetts: Heath-Lexington, Northeastern Studies in Vocational Rehabilitation No. 9, 1970.
93. Stromsdorfer, E.W., *Review and Synthesis of Cost-Effectiveness Studies of Vocational and Technical Education*. Columbus, Ohio: Ohio State University, ERIC Clearinghouse on Vocational and Technical Education, 1972.
94. Suchman, E., *Evaluative Research; Principles and Practices in Public Service and Social Action Programs*. New York: Russell Sage Foundation, 1967.
95. Thompson, J.D., *Organizations in Action*. New York: McGraw-Hill, 1967.
96. Trantnow, D., An introduction to evaluation: program effectiveness and community need, *Rehabilitation Literature*. 31 (1) (1970).
97. Tripodi, T., P. Fellin, and I. Epstein, *Social Program Evaluation: Guidelines for Health, Education, and Welfare Administration*. Itasca, Illinois: F.E. Peacock, 1971.
98. U.S. Department of Health, Education, and Welfare, Public Health Service, *Current Listing and Topical Index to the Vital and Health Statistics, 1962-1972*. Washington, D.C.: United States Government Printing Office, May, 1973.

99. University of Arkansas, *Arkansas Welfare Expansion Project: An Evaluation Proposal*. Fayetteville, Arkansas: Arkansas Research and Training Center, 1971.
100. University of Arkansas, *Analysis of Rehabilitation Counselor Sub Role Behavior: Development of a Rehabilitation Counseling Interview Sub Role Behavior Scale*, by S. Rubin. Fayetteville, Arkansas: Arkansas Rehabilitation Research and Training Center, 1973.
101. University of Minnesota, *The Measurement of Employment Satisfaction*, by R. Carlson, R.V. Dawis, G.W. England, and L.H. Lofquist, Minneapolis, Minnesota: Minnesota Studies in Vocational Rehabilitation XIII, Bulletin 35, May 1962.
102. University of Minnesota, *The Measurement of Employment Satisfactoriness*, by R. Carlson, R.V. Dawis, G.W. England, and L.H. Lofquist. Minneapolis, Minnesota: Minnesota Studies in Vocational Rehabilitation XIV, Bulletin 37, 1963.
103. University of Minnesota, *Manual for the Minnesota Satisfaction Scale*, by D. Weiss. Minneapolis, Minnesota: Minnesota Studies in Vocational Rehabilitation XXII, Bulletin 45, 1967.
104. University of Minnesota, *Manual for the Minnesota Satisfactoriness Scales*, by D.L. Gibson, D.J. Weiss, R. Dawis, and L.H. Lofquist. Minneapolis, Minnesota: Minnesota Studies in Vocational Rehabilitation XXVII, Bulletin 53, December, 1970.
105. University of Minnesota, *A Follow-Up Survey of Former Clients of the Minnesota Division of Vocational Rehabilitation*. Minneapolis, Minnesota: Minnesota Studies in Vocational Rehabilitation XXVI, Bulletin 50, 1969.
106. University of Oklahoma, *Management Practices in Vocational Rehabilitation District Offices*, by H. Vialle and W.G. Hills. Norman, Oklahoma: Regional Rehabilitation Research Institute, 1973.
107. University of Oklahoma, *Occasional Papers in Rehabilitation Administration; Third Party Agreements; Program Evaluation*. Norman, Oklahoma: Regional Rehabilitation Research Institute, 1971.
108. University of Wisconsin, Regional Rehabilitation Research Institute, *Correlates of Client Satisfaction in an Expanded Vocational Rehabilitation Program*, by K. Reagles, G. Wright and A. Butler. Madison, Wisconsin: Wisconsin Studies in Vocational Rehabilitation, Monograph XII, Series 2, 1970.

109. University of Wisconsin, Regional Rehabilitation Research Institute, *The Impact of an Expanded Vocational Rehabilitation Program Upon Intra-Agency Processes and Procedures*, by S. Katz, G. Wright, and K. Reagles. Madison, Wisconsin: Wisconsin Studies in Vocational Rehabilitation, Monograph XIV, Series 2, 1971.
110. University of Wisconsin, Regional Rehabilitation Research Institute, *Rehabilitation Client Sustention: A Longitudinal Study*, by D. Gay. Madison, Wisconsin: Wisconsin Studies in Vocational Rehabilitation, Monograph XVI, Series 2, 1971.
111. University of Wisconsin, Regional Rehabilitation Research Institute, *A Scale of Rehabilitation Gain for Clients of an Expanded Vocational Rehabilitation Program*, by K. Reagles, G. Wright, and A. Butler. Madison, Wisconsin: Wisconsin Studies in Vocational Rehabilitation, Monograph XIII, Series 2, 1970.
112. Van Kruegel, E., Introduction to CPM, *The Architectural Record* 74 (9) (September, 1964).
113. Walker, R., The ninth panacea: program evaluation, *Evaluation* 1 (1) (Fall, 1972).
114. Washington, Vocational Rehabilitation Division, *Evaluation and Inventory Project; Final Report*, by R. Schumann. Olympia, Washington: 25-P-65216/0-03, 1972.
115. *Webster's Seventh New Collegiate Dictionary*. Springfield, Massachusetts: G.C. Merriam, 1971.
116. Wheeler, D., *The Measure of Job Relatedness for Occupational Evaluation*. St. Paul, Minnesota: Minnesota Research Coordinating Unit in Occupational Education, 1971.
117. Wholey, J., et al., *Federal Evaluation Policy: Analyzing the Effect of Public Programs*. Washington, D.C.: The Urban Institute, 1970.
118. Wright, G., K. Reagles and K. Thomas, The Wood County project: an expanded rehabilitation program for the vocationally handicapped, *International Labor Review* 104 (1-2) (July-August, 1971).
119. University of North Carolina, School of Public Health, *Some Notes on Statistical Technique*. Chapel Hill, North Carolina: mimeo, 1967.

Appendix I

Other Classes of Measures Used in Program Evaluation

In addition to the outcome measures of effort, effectiveness and efficiency some authors have defined other similar and overlapping measures. These are:

Accomplishment—program outcome (effectiveness). (60)

Adequacy of Performance—degree to which the performance of the program meets the total amount of need. It answers the question "How effective is it compared to need?" The unique factor in Adequacy of Performance is the comparison to estimation of need or some other standard. (94)

Performance—program outcome (effectiveness). (94)

Process—the detailed determination of activities undertaken to affect program outcome. (94)

Productivity—loosely interpreted to be the efficiency with which output is produced by the resources utilized. A common productivity measure is:
$$\frac{\text{output}}{\text{man hours}}$$

Below are listed four classes of measures popular in process evaluations.

Accessibility—judgmental statement concerning the problems of entry into the system. It answers the question—"Can a client have services at a time (hours of the day) and place where he needs them?" (2)

Availability—judgmental statement concerning the supply of needed services. It answers the question—"Do the necessary services exist in this locale?"

Comprehensiveness—judgmental statement concerning the range of available services. "It answers the question "Will the client receive all needed services in 'one stop'?" (2)

Continuity—judgmental statement concerning the program's responsibility for maintaining an ongoing relationship with the client, and the coordination of services with other community services. It answers the questions "Is this a fly-by-night operation?" or "Do we feel a responsibility to follow-up on clients?" (2)

Appendix II

Standards

As stated in Chapter I, standards of program performance may be set from indices. There are several characteristics which usually qualify indices for use as standards of program performance and professional practice. Included are consensus, stability, validity, and transferability, level of stringency, explicitness and content.

Consensus—The standard should be set with a certain degree of agreement among recognized leaders in the profession.

Stability—Since standards are influenced by technologic and philosophical change, usually there is a need to reformulate a standard constantly.

Validity—The standard must show a demonstrated relationship to the program outcome. In other words, the standard should measure the variable that explicitly represents effectiveness, (e.g., the number of clients employed competitively).

Transferability—The ability to apply the standard to more than one group and make comparisons, (e.g., from state to state, or from generic program to generic program. However, some variability in standards can be appropriate under special local or program conditions.)

Level of Stringency—Indicates the degree to which the standard is binding or strictness to which it is enforced. In addition, the level of stringency is also taken to mean how high the standard is set. If the standard is set very high, few will meet it; if set very low, all will meet its requirements. The standard can be set as a minimal standard, an optimal standard or an "average" standard. Each has its particular difficulties.

Explicitness—Preciseness of the criteria used as a standard. If possible, the standard should be so precise as to leave no doubt as to whether the standard has been met in any given situation.

Content—Some standards use many criteria to establish what programs are good, or a standard may employ only one or two criteria for goodness. Content refers to the identification of the criteria of goodness included in the standard so that the range of possible criteria believed necessary to determine "goodness" is covered. (30)