

UC Berkeley
IURD Working Paper Series

Title

An Overview of Tools and Requirements for Program Evaluation in Vocational Rehabilitation Agencies

Permalink

<https://escholarship.org/uc/item/1v76b96m>

Author

Collignon, Frederick C.

Publication Date

1973-11-01

AN OVERVIEW OF TOOLS AND REQUIREMENTS
FOR PROGRAM EVALUATION
IN VOCATIONAL REHABILITATION AGENCIES

Frederick C. Collignon

November 1973

Working Paper No. 248/RS024

Frederick C. Collignon, Project Director - Michael B. Teitz, Principal Investigator

Project for Cost Benefit Analysis and Evaluation of Rehabilitation Services

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

PREFACE

This paper was initially presented at the annual convention of the National Rehabilitation Association in Atlantic City, New Jersey, November, 1973, at the first symposium given by N.R.A. on program evaluation. The symposium was sponsored formally by the Administrative and Supervisory Practices Division of N.R.A., which subsequently reproduced and circulated the papers presented to Division members nationally. Because of the many requests for the papers, their printing was exhausted. With ASPD encouragement, I am thus making the paper available through our Institute series.

Frederick C. Collignon

Introduction

The 1970's have been widely hailed or derided, depending on the audience, as a decade of accountability. The share of the national economy or G.N.P. being produced by government has risen dramatically over recent decades, and taxes have increased accordingly. In the private sector, managers have profits as indicators and incentives for success and potential profits as criteria for making decisions. In the public sector, there are no profits. Indicators and incentives for success have traditionally been a growing program budget and signs of approval from legislatures and a Governor's Finance Department. Whether such success indicators have had any link to program performance in meeting publicly pronounced program objectives has been dubious. In the decade of accountability, if one is to believe the popular pundits, the taxpayers, the consumers of services, and their legislative representatives have said "Enough!" to growing public budgets and taxes, and have demanded to know what is being received for the money expended. Thus, the new thrust is for clearer goal definitions for programs, the careful specification and measurement of indicators of program output and impact, a greater consciousness of the need to justify the inference of causality between program activities and observed changes in behavior or situations, the requirement that benefits be evaluated with reference to the costs of achieving the benefits, and finally the insistence that alternative strategies for achieving goals more efficiently be compared. Whether the

demand for program evaluation represents merely the latest in the line of panaceas for making bureaucracies perform, a line that runs through PPBS, Management By Objectives, and numerous other innovations in public sector management remains to be seen.

I. Evaluation - Organizational Challenge and Source of Confusion

Given the context of this call for accountability, public agencies and rehabilitation agencies in particular have begun placing greater stress on evaluation activities. The new legislation demands evaluation. Most state agencies within the last three years have assigned one or more staff with the formal responsibility for evaluation. Over the last five years, first R.S.A. and more recently S.R.S. have begun to include program evaluation and policy analysis within their research priorities.

A review of the literature in the rehabilitation field and a recently completed survey of the activities of state rehabilitation agencies suggest that much basic work remains to be done. The available examples in the rehabilitation literature of attempts at program evaluation are few. It was striking that the otherwise excellent review article by Hefferlin and Katz in Rehabilitation Literature two years ago, which summarized a large number of evaluation studies in the rehabilitation field, failed to include any examples where the effectiveness of services was compared to the cost of services, or where costs or the worth of client gains were even considered at length.

¹ Elizabeth A. Hefferlin and Alfred H. Katz, "Issues and Orientations in the Evaluation of Rehabilitation Programs: A Review Article," Rehabilitation Literature, 32, Part I (March, 1971), 66-73, and Part II (April, 1971), 98-106.

The evaluation studies reported may have occasionally been good applications of experimental design, but they were clearly limited as program evaluations. Indeed, one of the problems is that most program evaluation studies in the rehabilitation field have been in the form of internal or contracted studies conducted by state and Federal agencies. Even when of high research quality, such studies often do not receive wide circulation nor are submitted to journals for publication, in part because they might prove controversial, in part because they are directed at particular decisions or probe particular agency problems which might not be of interest to journal audiences, in part because the methodology does not conform to normal research standards. When published, evaluation studies are more likely to appear in the kinds of journals (e.g., economics, public administration, planning, operations research) which are not widely read by rehabilitation program managers and evaluators who have come up from the counselor ranks and who are more commonly disposed by professional training to the journals of counseling, health, and education.

State rehabilitation agencies, although now widely engaged in activities they call evaluation, widely vary in their activities.¹ Most state agencies have small evaluation units often with multiple responsibilities in addition to evaluation and usually staffed pri -

¹ See the analysis of a national survey in Susan Shea Ridge, Survey of State Program Evaluation Practices, Institute of Urban and Regional Development, University of California, Berkeley, April, 1973, Working Paper No. 209/RS20.

marily with former rehabilitation counselors promoted through the ranks under the aegis of civil service. These agencies are often the first to admit their confusion in understanding what evaluation is supposed to accomplish and - more importantly - how they are supposed to proceed to carry out a "good evaluation." Only the larger states have shown much success in launching a comprehensive evaluation program of competently executed studies which prove effective in providing assistance to state directors in improving program management. This success for a few large states is attributable - I believe - to the greater amount of funds for staff and analysis available in those state agencies and to the ability or foresight in those state agencies to hire staff with existing skills or to train existing staff in the skills necessary for program analysis and evaluation: public administration, economics, statistics, and methodology and research design.

In a rather long paper given at the ARCA convention in San Diego last February¹, I discussed the organizational conflicts over evaluation that have arisen within state agencies, between Federal and state agencies, between rehabilitation agencies and their legislatures and monitoring offices in the Executive branch of their state governments, and finally between the newly militant consumer organizations of the disabled and rehabilitation agencies and professionals.

¹Frederick C. Collignon, An Overview of Program Evaluation Activity in Rehabilitation Services Programs: Current Status and the Problems Ahead, Institute of Urban and Regional Development, University of California, Berkeley, February, 1973, Working Paper No. 207/RS019.

I also reviewed my experience with the Institute on Rehabilitation Services in 1971-72 which produced the excellent introduction to program evaluation for state agency staff entitled Program Evaluation: A Beginning Statement¹. I do not want to repeat that discussion here, but I would like to emphasize again several major points from that paper.

First, there are many differences among those calling for evaluation in their concept of why evaluation is important and what evaluation should be. These differences give rise to very different types of activity within evaluation units. The individual who perceives evaluation as an activity for holding a program accountable, or for improving program management and performance, or for spurring innovation, or for research, tends to use different methodological approaches, emphasize different criteria for program success, assume different roles as evaluators, and look to different audiences inside and outside the program for approval.

Second, reflecting these differences, there is no single definition of "evaluation" or set of methodological rules for evaluation which can be given to the staff of a small state agency trying to develop an evaluation component. Proponents of experimental design will immediately denounce such a statement, and I am sympathetic to

¹ Tenth Institute on Rehabilitation Services, Program Evaluation: A Beginning Statement (Social and Rehabilitation Service, U.S. Department of Health, Education and Welfare, May, 1972). The report is also available from the West Virginia Research and Training Center, which sponsored the I.R.S. Study Group.

the assertion that experimental design represents a model by which much evaluation study should be in turn evaluated as research. Accepting experimental design as the standard for good evaluation presumes, however, that evaluation is primarily concerned with scientific documentation of the causal relationship of program intervention to client outcome. Unfortunately, evaluation has broader purposes, including providing help and insight in improving planning and management, assessing alternatives, making specific decisions, and implementing program change. For these purposes, experimental design is at best of limited use, if not irrelevant. Evaluation when used for these purposes is an activity which uses methodology from a number of different social science disciplines in a highly eclectic and imperfect manner.¹

Third, much of the confusion and conflict which arises within organizations over evaluation stems in part from a failure to distinguish between the external and internal functions which evaluation can perform.

Agencies often launch evaluation units not because they believe

¹The field of evaluation, to the extent that there is a unique intellectual field with a peculiarly characteristic collection of techniques and problems requiring solution is roughly at the stage of development where operations research was some fifteen years ago. Indeed, the danger is that, like operations research, evaluation will increasingly receive attention in the university, become an identified field with its own institutional base, and refine its methodological tools, while at the same time becoming esoteric in application. Most evaluations of the utilization of operations research findings have been negative. Evaluations of evaluation to date have similarly been negative, although it can be argued that such an assessment is still premature. Cf. Stimson,

that they will learn better ways to manage their programs or ways to increase goal achievement, but because they are obliged to evaluate programs by Governor's offices, legislatures, or Federal funding agencies in the region or Washington. Evaluation activity becomes a form of "external program defense" from the perspective of the agency. Evaluation reports are seen as necessary to insure continued funding or at least to get outsiders "off the back of the agency". In such situations, the evaluation unit functions as a public relations arm of the agency. As with any public relations activity, there is much incentive to report those aspects of program operations which are favorable, to overlook aspects which are unfavorable, and to manipulate information for the purposes of protecting the program. The evaluation unit is kept honest by the extent to which its statements are independently audited by outside monitors. Friendly monitors are always in high demand. The parallel between, first, the behavior of counselors when evaluated by central office and, second, the central office when evaluating itself or being evaluated for the sake of outside agencies is clear. This program defense function of evaluation activities is the opposite side of the coin from the function of accountability. Few state directors and program managers have been deceived by high-minded calls for accountability; like their counselors, they recognize the potential threat in being evaluated by outsiders.

Unfortunately, the need to use evaluation for program defense often detracts from the use of evaluation for improving internal management and planning. The problem has two aspects.

First, outsiders and insiders want the answers to different questions. Outsiders allocate limited budgets between rehabilitation and other goals and have to choose between different program strategies (e.g., manpower training, integrated social services, basic income support coupled with service vouchers, the "vocational rehabilitation approach") for achieving social objectives in dealing with the disabled. For such tasks, overall program benefit-cost ratios may have some use, as may knowledge concerning whether in fact the rehabilitation program prompted a change in client employment which would not have been achieved in the absence of services. Knowledge about the effectiveness of program subcomponents (one office versus another, alternative service mixes) will have less use, because the Governor's office and legislature lack the time and energy, if not the interest, in trying to delve in to the detailed management problems of a program as small and as (normally) noncontroversial as vocational rehabilitation. For the state director, whose program must function within defined legislative boundaries and who lacks the authority often to adopt major new technological alternatives to the traditional rehabilitation approach, information about the overall program benefit-cost ratio and the net gains made by clients served over some control group is not particularly helpful in making decisions. More useful in providing information on how to improve performance would be evaluations of program subcomponents.¹ In short, evaluation assists in

¹ Similarly, an evaluation of program subcomponents is usually at too high a level of aggregation of program operations to be of use to the counselor. He may be affected by a decision to increase or reduce the fund-

improving management to the extent that it provides information relevant to the decisions and actions which the manager has the responsibility to undertake. Given limited funds for evaluation activity, however, the program defense functions of evaluation often receive first priority and few resources remain for directing evaluation at internal management needs.¹

The second aspect of the conflict between the program defense and internal management functions of evaluation centers on the different kinds of indicators which are appropriate for such purposes.

Looking to those outside the agency, the state director normally wants measures of success which are simple to explain, which will tend to show an improvement in program achievement over the previous year, and which appear to show significant social value being generated by the program. Benefit-cost ratios and the number of 26

ing of his office or component of the program, but he cannot use the information provided by such an evaluation in his own decision-making. Even an evaluation of workshops versus other kinds of training is only useful to the counselor if he can infer that the workshops and vendors in his particular area are indeed comparable to the "average" workshop and training source that emerge in the sample or universe studied in the evaluation.

¹ Why program defense receives first priority needs explanation. The threat of the withdrawal of funding and political support by outsiders obviously creates incentives to provide the evaluations demanded. It is unclear from where the incentive comes for using evaluation to improve internal management and program performance. There are no profits to create incentive, nor is organizational survival dependent on improved performance. Programs once created rarely are disbanded. We must look instead to some sense of professional integrity to explain why rehabilitation program managers might take the initiative in seeking ways to better satisfy the needs of taxpayers and the disabled at less cost.

closures may often appear to be excellent indicators of program success from the perspective of the director, in part because they tend to show the program in a flattering light. The number of 26 closures can easily be manipulated by shifting the mix of clients accepted, by increasing the number of homemaker closures, by ignoring the probable stability and career potential of the jobs to which clients are referred, and so on. Benefit-cost ratios tend even when sensibly estimated to be very favorable to vocational rehabilitation programs. They also are easily manipulated, however, by choosing low or no discount rates, ignoring follow-up data on the stability of jobs, ignoring costs borne by the client and other agencies, treating any wages received by 28 closures as benefits attributable to rehabilitation services, reducing the time period prior to program referral which is used for measuring the employment potential of the client in the absence of services, and so on. Program analysts outside the agency may be insightful to catch some of these manipulations, but seldom have time - given the obligation to review many programs in addition to rehabilitation - to become familiar enough with the program to contest the assumptions behind the indicators.

The problem, however, is that the use of such indicators for external defense creates future external demands and thus internal incentives to increase performance in terms of those indicators. He who lives by the numbers games shall die by the numbers games. Even worse, the need to generate such indicators for outsiders also often overwhelms the information system of the agency, confuses some central

office analysts into believing that such indicators are sufficient for internal management purposes, and produces wholesale cynicism within the agency about evaluation. It has generally been my experience that counselors and program analysts inside rehabilitation agencies understand better the pitfalls of using the 26 closure to evaluate rehabilitation services, than do legislators, Finance Department analysts, and even client lobbies. Yet many agencies continue to use the number of 26 closures, and the percentage of accepted clients closed as 26's, and the cost per rehabilitation as the principal indicators of success in evaluating counselors and program components. Better indicators, even when routinely measurable now with existing R-300 data - such as the sufficiency of the job's wage to lift the client's family out of poverty, the increase in earnings generated by a dollar's expenditure for services, and the ratio of actual to expected rehabilitation success and costs given caseload composition - are ignored.

II. Evaluation - What is It?

Having given some image of the varying goals which are set for evaluation and of the organizational problems and priorities likely to confront the program evaluation unit within the state agency, let me briefly provide an overview of the kinds of evaluation activities taking place in state agencies.

Under the rubric of program evaluation in rehabilitation programs, three different activities are undertaken, each of which sheds considerable light on current program performance and possible di -

rections for improvement. The three activities are case review, case flow analysis, and overall program impact assessment. Agencies which have been engaged in evaluation activity for many years and which have large and sophisticated evaluation staffs often lodge the case review and sometimes the case flow analysis activities with operating divisions. They view those activities as essentially management and monitoring rather than evaluation functions. In contrast, agencies just beginning to develop evaluation capability - the usual situation for most state rehabilitation agencies - often concentrate most of their resources on case review activities. Only as the agencies gain more experience and sophistication do they perceive the need and develop the capability for first, case flow analysis, and then much later, overall program impact assessment.

The guidelines on evaluation prepared by the Institute on Rehabilitation Services suggest five broad types of criteria for evaluating programs: client and community impact, higher-order program efficiency, program effectiveness, lower-order program efficiency, and program management.

Case Review is directed at evaluating program performance in terms of program management criteria. Individual cases are reviewed in terms of conformity of counselor actions to standards

for professional practice set out in state manuals or policy pronouncements. The actual impact of services upon the client is not examined, except perhaps to note whether the closure status of the case appeared appropriate in view of the recorded information on the change in the client's situation.

Case flow analysis is directed at evaluating the program in terms of program input (e.g., the turnover of clients within various statuses, the overall balance among statuses within caseloads, costs of services per status change or over some unit of time). The criteria are those of lower-order program efficiency.

Overall program impact assessment is directed at the other three types of criteria. This type of activity tries to analyze program output and relate output to program input.

At the simplest level, most state agencies monitor the total number of rehabilitations per year, per counselor, and per 100,000 base population. These measurements represent an evaluation of program effectiveness. Unfortunately, such evaluation accepts at face value the validity of the 26 closure status as a measure of rehabilitation and thus program impact, even though the placement of a client in that status is based often on highly judgmental and arbitrary decisions by counselors. Moreover, the 26 closure measure fails to provide information on the quality of

rehabilitation or the magnitude of improvement of the client's situation. The measure, unless adjusted, also fails to distinguish the relative difficulty of cases.

The evaluation of the program in terms of client and community impact represents a dramatic leap in the quality and sophistication of an agency's program evaluation activity. The actual impact of the program upon the client which is hidden behind the 26 closure status becomes the focus of attention. The change in earnings, occupation, and other client functional capabilities receives much attention. Client evaluation of the usefulness and quality of services received may be directly solicited. The stability of program impact over time may be explored by follow-up studies of clients whose cases had been closed several years. Such studies may reveal increases in client earnings over the earnings level at closure which are attributable to program impact. Such studies may also reveal, however, that some clients have dropped out of employment or suffered a deterioration in homemaking capability due to worsening of disability, shifts in client motivation and morale, and perhaps the insufficiency and poor quality of the rehabilitation services received earlier.

The final phase of program impact assessment activity is the evaluation of the program in terms of higher-order efficiency. Only in this phase are program outputs directly related to program inputs. At the simplest level, state agencies have routinely measured costs per

26 closure. This efficiency measure again suffers the limitation of accepting the closure status at face value. As agencies advance in sophistication, they may begin to substitute change in earnings for the 26 closure status as the output proxy in efficiency analysis. The program is then evaluated in terms of dollars of increased client earnings per dollar of program costs. Evaluation in terms of this criterion resembles, but is not the same as, benefit-cost analysis, which represents the most advanced form of evaluation currently practiced. Benefit-cost analysis looks at comprehensive social benefits rather than simply change in client earnings. Life-time earnings changes are also examined rather than simply the observed immediate change in earnings. Also, social costs are analyzed rather than solely the costs borne directly by the rehabilitation program. The relationship of benefits to costs, whether expressed in terms of benefit-cost ratios, internal rates of return, payback periods, or net present values, usually becomes the ultimate criterion for evaluating the program in this type of evaluation, although there is no necessary reason why the various valued benefits and costs (some of which may be difficult to measure or even inherently intangible) need be reduced to a single aggregated number.

There is a further phase of program impact assessment which currently is not pursued by any state agency, but which is fundamentally important: exploration of the causal role of program services via experimental design and the use of control groups. This type of evaluation, which university researchers often consider the only type of

valid evaluation, is called "evaluative research" because of its use of "scientific" method and basic research techniques. The operational difficulty of assembling experimental and control groups, the questionable morality and the political difficulty in denying control groups services, and the costs of such evaluation studies have discouraged state agencies, the Federal government, and university researchers from undertaking this phase of evaluation. Thus, the question of what would have happened to clients in the absence of services, remains unanswered. The extent to which rehabilitation services actually "caused" the improvement in the client's situation is not known. Pessimists and cynics might speculate that all observed improvements would have occurred anyway, even without services. Those more sympathetic to the program maintain that the improvement observable in a before-after comparison of the client's situation is directly (causally) attributable to program intervention and the receipt of services.

Although most state agencies in the rehabilitation field are just beginning to move into the area of program evaluation, several of the more advanced states are already experiencing a problem which has been encountered in other program areas which much earlier moved into evaluation (e.g., education, poverty programs, health). This problem is essentially that of implementing change based upon evaluation findings. It is one thing to establish that current program performance is not up to desired standards, it is another thing to determine why, and it is still another challenge to be able to change

behavior and improve program performance. The frustrating experience in many programs has been that evaluation studies remain "on the shelf" and do not lead to policy initiatives, revised resource allocation, and decisions. Even when evaluation studies have produced decisions and actions directed at producing change, change has been hard to achieve. The implementation of change has emerged as a major problem in its own right. Only recently and not in the rehabilitation field, have university researchers begun to become aware of the problems concerning the implementation of change. Much attention is being given to the process of evaluation and the development of evaluation systems which stress process as much as the collection and analysis of data.¹ Unfortunately, although much theory has emerged, clear and proven guidelines for program action do not yet exist. Demonstrations will be needed before theory can be verified and guidelines developed.

III. Evaluation Tools - Organizational and Functional Requirements

As my introduction suggests, there is no single right methodology for evaluation or for analysis of a management problem. In this section, however, I would rather audaciously like to discuss some basic requirements for an effective program evaluation unit within a state

¹ Cf. Glenn Siebert, Implementation of Evaluation and the Systems Approach in Government: A Literature Survey and Conceptual Model, Institute of Urban and Regional Development, University of California, Berkeley, January, 1973, Working Paper No. 201/RS016.

vocational rehabilitation agency. Then I would like briefly to review some of the uses and misuses of the more commonly employed alternative evaluation methodologies.

Among the organizational and functional requirements for an effective evaluation unit which I would list are the following:

1. Staff Resources. The most important resources available to an evaluation unit are the methodological skills, theoretic knowledge, understanding of organizational behavior, ability to conceptualize policy and management questions, and knowledge of rehabilitation programs possessed by the staff themselves. In-service training in evaluation methodology is usually insufficient to compensate for the lack of these basic skills (called "human capital" by some). Nor can the employment of consultants or linkage of staff with university groups provide a workable substitute. The tendency in most states nonetheless is to fill evaluation units almost completely with former rehabilitation counselors who have come up through the ranks, and to rely on in-service training to provide evaluation skills. Civil service regulations often encourage such promotions, and there is the belief among some state directors that such individuals better "understand" the rehabilitation program they are evaluating. Generally, however, such staff lack the skills and background for program analysis. There are exceptional individuals who succeed in making the transformation from counselor to program analyst/evaluator, but closer checks usually reveal that such individuals had strong

research interests and some relevant training prior to their counseling careers. Generally, state agencies would be well advised to man most of the positions in evaluation units with individuals specifically trained in methodology, economics, public policy, public administration, and analysis. In other programs, one normally has to counsel program directors to include at least one person with operational experience in the evaluation unit. In the rehabilitation field, the reverse is true: most evaluation staff have operational experience, but few staff have any skills relevant to program analysis and evaluation.

2. Staff Time. A second key resource to be furnished any evaluation unit if it is to be effective is sufficient time to engage in evaluation. When evaluation staffs are good, there is strong incentive to give them many additional functions in helping the agency put out the innumerable fires which are routinely igniting. The evaluation staff tends to find more and more of its time being spent in planning, task force work, and similar activities which while very useful to the agency (perhaps more useful in some cases than evaluation), prevent the unit from engaging in evaluation.

3. Computer Access. Another key resource for any evaluation unit is access to some computer facility with at least minimal software social science packages for data analysis. I have seen agency staff spend six man-months of labor in laboriously calculating statistics by hand or desk-calculator which could have been generated in a few days time at most by an analyst having access to a computer

programmer. Given the increasing size of the rehabilitation programs, data processing capabilities are simply critical to an evaluation unit. It must be stressed that the existence of a management information system linked to a computer is not sufficient for assuring access of an evaluation unit to computer capabilities. Many state agencies, large and small, have "information systems" which have gotten out of control, generating inaccurate data at time intervals far too frequent to allow absorption by program managers. Often the agency's computer time allowance is completely captured by the ongoing needs of the information system. There is little time remaining for analysis or special data processing to meet the needs of evaluation.

4. Organizational Contact with the Director. If an evaluation unit is to have any impact on program decisions, it must have contact to the Director to know what decisions are pending and require information. Similarly, only contact with the Director allows a unit to identify what are pressing problems requiring analysis and solution. Moreover, if the Director is to become aware of and understand the findings of evaluation studies, issue memoranda and personal briefings are likely to be more effective than long reports. Finally, the evaluation unit is much more likely to receive the cooperation of the line units within the program if those units are aware that the evaluation unit is acting at the directive of the Director, and that the Director takes a personal interest in the work of the unit.¹

¹All state directors need, of course, is yet another unit directly

5. Broad Participation in the Evaluation Process. One of the few general conclusions of the literature on the utilization of evaluation studies is that utilization is more likely, the broader the participation by program managers and line staff in the process of planning, conducting, and assessing the study. Broad participation allows evaluation staff to gain better insight into the program being evaluated and thus to be better able to state why program performance may not be up to desired standards, if such findings do emerge. Similarly, participation results in more suggestions to the evaluation unit concerning how performance could be improved. Participation also increases the understanding of line staff as to why change may be necessary and what kinds of change are needed. As a result, their resistance to innovation and change is reduced.

Such a call for broad participation does not imply that the operating staff themselves should conduct the evaluation. The skills of the evaluation unit are still required for study design and implementation, data analysis, and report-writing. Evaluation units should not function as isolated research entities, however. For given studies, advisory groups of counselors, administrators, and even clients are usually desirable. The evaluation unit should also solicit routinely suggestions from the line units and from the field on problems requiring analysis. The unit might thus serve as a two-way information channel

reporting to them. My emphasis here is on the use of evaluation in shaping decisions, i.e., internal management and improving performance. When the functions being performed by the evaluation unit fall into the domains of program defense or internal accountability, there is no particular reason why direct contact with the director is necessary.

between operating staff and the Director, paralleling regular line channels, but focusing more on generic problems impeding program performance.

6. Clarity on Goal Conflicts. Much of the evaluation literature asserts that evaluation can only be successful if clear goal definitions and priorities have first been given. I personally agree that evaluations would be much easier if such goal definitions were available, but I don't believe that such definitions are necessary for effective evaluation. Public sector organizations resist goal setting at other than the most abstract levels. Priorities are even more resisted. This is not due so much to bad management, as to incentives built into the political process within which most public sector programs operate. Legislation is usually generated by log-rolling, or coalition-building, such that a program is sold by appealing to many different constituencies, often with differing and even conflicting interests and priorities. Any formal statement of goals and priorities might threaten the coalitions upon which secure program funding rest. An attempt to establish priorities in rehabilitation service programs, for example, might well result in pitting various disability groups explicitly against one another in seeking a larger share of the limited program budget. Such competition might well undermine the needed coalition for securing a larger overall budget so that more of the unmet need in the population might be addressed. Similarly, within a program system at any given time, there are many different people with varying goals and commitments. Goal consensus is impossible.

Nor can evaluators expect that an agency director's priorities will receive unquestioned acceptance from professionals within service programs like vocational rehabilitation; professionals prize too much their autonomy and in few state programs is the director really sufficiently powerful to impose his will to bring about uniform and quick changes in behavior. As Chester Barnard wrote in the 1930's, the key function of the executive is persuading people to act in ways that achieve organizational goals.¹ Orders don't simply flow through the bureaucracy and become implemented because they were issued by the director.²

What I believe important then for an evaluation unit is not clear goal definitions, but rather a clear awareness of the goal conflicts among different people and different levels within the rehabilitation system. Such an awareness proves useful in deciding what questions to pursue in the evaluation and in formulating recommendations following an evaluation. The evaluation unit must become accustomed to working in a situation of ambiguity, where the evaluation unit is helping the agency define and even discover its operational goals via evaluation, rather than evaluating in terms of pre-given goals. I stress "discovery", because I believe that one of the more

¹ Chester I. Barnard, Functions of the Executive (Cambridge, Mass.: Harvard University Press, 1938).

² I am aware from agency friends across the country that in some states, especially those in the South, the older image of the state director as all-powerful executive remains strongly held and may even correspond to reality in a few cases.

useful functions of evaluation is the identification of those goals which are implicit in the decisions and activities of the agency. Often it is only when the agency is confronted directly with the goals or priorities which are implicit in organizational behavior that a change in priorities or more explicit goal-directed behavior will be forthcoming. In the attempt to avoid the stress of making goals explicit, individuals forget that there are implicit tradeoffs in any action they take.

7. Use and Understanding of Multiple Indicators. Evaluation units can best educate the agency in goal-setting by insisting on using a wide range of indicators that address and clarify achievement of different supplementary goals which most individuals in the organization would espouse. Because most programs have multiple goals in the public sector and because many of the important goals are not readily translated into dollar benefit terms, evaluation units should insist on using multiple indicators. The use of multiple indicators can also assist in program defense, since performance seldom is bad in terms of all indicators; the agency can select from a number of alternative measures of success in depicting its performance to outsiders.

Different indicators reflect different goals. Let us consider several indicators currently available for use by rehabilitation agencies. The 26 closure, to the extent that such a closure does in fact reflect a major change in an individual's situation such that the individual is now effectively performing as a homemaker or employed

in a job, is a reasonable indicator if the purpose of rehabilitation services is primarily to maximize the number of people moving into productive activity of whatever kind. In this society, engagement in productive activity conveys dignity, status, and self-worth. Thus, the engagement of people in productive activity is a meaningful social goal independently of the income consequences of such activity. In fact, we are all aware of the deficiencies in the 26 closure as used in practice as a measure of placement in more productive activity. Yet, if agencies were to audit cases carefully, it is conceivable that the 26 closure might be a good proxy for rehabilitation.¹ The cost per 26 closure is a measure of the efficiency of the program in shifting people into productive activity. The ratio of 26 closures to all cases accepted is a measure of effectiveness relative to effort in achieving the movement of people into productive activity. Similarly, one can develop intermediate measures of success (size or ratios of different closure statuses) which help monitor the flow of individuals through the system. For all these indicators, their linkage to the 26 closure, given the current criteria for such a closure, reinforces the role of rehabilitation programs in providing people with productive work and employers and families with productive workers. Nothing in the indicator directly reflects, however, the degree of increase in productive activity, the extent to which a

¹What would still be needed, even if such audits were routine, however, would be some measurable indicators of what improved homemaking capability entailed. Unlike job placements, closures in homemaker statuses are not readily capable of being evaluated by auditors (or even by the counselors themselves.)

permanent change in the individual's situation has been achieved, or the returns to societal investment in the program.

If the goal of rehabilitation programs were to assist individuals in escaping poverty and dependency upon others, very different kinds of indicators would be appropriate. Agencies might look at change in the amount of welfare received by the individual or whether the individual has gone off the welfare rolls. Because rehabilitation services are not particularly successful in moving people off the welfare rolls altogether, these kinds of indicators are seldom emphasized because they look bad in defending programs. What many agencies neglect, however, is understanding why so little change in welfare status occurs. A useful indicator here would be the ratio of the wage at closure to the minimum wage. Such an indicator would suggest whether or not the degree of vocational improvement being experienced by clients was likely to be sufficient in removing them from welfare rolls. Weighting the state minimum wage by the ratio of the median state income to the median national income would be useful here as well, to control for the widely varying costs of living among the states. Receiving the minimum wage in a Southern state implies a much higher standard of living and vocational achievement for a disabled individual than receiving the minimum wage in a large city or Eastern or Great Lakes state.

Generally, greater attention to the level of the wage received at closure gives implicit priority to the goal of helping the individual achieve permanent vocational self-sufficiency. The labor market

economists tell us that low wage jobs tend often to fall into the secondary labor market and to be characterized by frequent turnover, few fringe benefits (now more than 20% of the total payrolls of American manufacturers), and little opportunity for training, skill-acquisition, or internal promotion. The wage level at closure and the change in wage level from referral to closure measure the actual improvement in the individual's situation. They could be regarded as good measures of the "quality of service", a frequent concern of rehabilitation professionals. They also serve as good proxies for the stability of employment after the closure is recorded. Unfortunately, the change in wage levels is less useful for evaluating the services received by some individuals than is the wage level at closure, because the R-300 system measures wages only in the week prior to the referral when the individual is usually unemployed. In order for the change in wage levels to be meaningful, we would want to know the wage level enjoyed by such individuals in their last employment prior to referral, but following disability. The portion of unemployed referrals for whom continued unemployment is the probably situation in the absence of services is unknown, but is certainly much smaller than is suggested by the R-300 record.

The ratio of wages at closure or the increase in wages to the costs of services addresses as an indicator goals of efficiently increasing the total productive activity of the society. The focus is not on the number of productive people but rather the amount of productive activity. Such indicators are proxies for the return on

societal investment. The indicator can be modified to reflect the return on the taxpayer's investment by substituting in the numerator of the ratio; the change in taxes paid and in public assistance payments received by the client.

Other indicators are needed to reflect different goals. The search for "valid" and statistically "reliable" measures of independent living reflects a concern for the impact of services on client's lives that complement or transcend changes in vocational activity. The use of consumer evaluation and the measurement of the satisfaction of clients with services received addresses the goals that services should address client wants and needs as well as the taxpayer's interests and that the process of service delivery should respect the dignity and feelings of the clients.

In using any indicator, the comparison of the indicator to the costs of achieving the level of the indicator allows the agency to consider the efficiency of its efforts in achieving success. The justification for spending \$10,000 or more for an individual case on the grounds that it is impossible to measure the value of improving the quality of an individual's life or even maintaining life strikes the heart-strings. Yet implicit in such a decision is the foregoing of opportunities for having a possible profound impact on several other lives.¹

¹ It may be possible in many cases (e.g., dialysis machines) to justify such large expenditures by the public sector in terms of the "option value" of maintaining the access of all citizens to such services if

Another general suggestion that applies to almost any indicator that is used is that aggregations of indicators for rehabilitation programs should be adjusted to reflect case mix. Adding the achievements of services with different disability groups or with individuals with varying educations or supplemental characteristics creating difficulties in the labor market (old or very young age, sex, minority status) is an addition of apples and oranges, given the purpose of evaluating performance. A simple weighting of the actual performance with the case load by the expected performance given the case-load's characteristics is one way to prevent "creaming" strategies from distorting aggregated indicators.¹

they become disabled. Ronald Loshin and I have a paper forthcoming estimating the "option value" of the Federal rehabilitation program.

¹Of course it is still necessary to decide what are the most important characteristics affecting a case's difficulty. For most purposes, a control for disability may be sufficient. A weighting factor by disability, age, sex, race, and education is not particularly difficult, if the agency has access to a computer. Cf. Philip A. Armstrong, Program Analysis: Patterns of Cost, Output, and Productivity Among Districts in a State Rehabilitation Program, Institute of Urban and Regional Development, University of California, Berkeley, November, 1972, Working Paper No. 204/RS018; and David E. Serot, Indices of Cost, Output and Productivity for Use in Evaluating Rehabilitation Services Programs, Institute of Urban and Regional Development, University of California, Berkeley, August, 1972, Working Paper No. 187/RS013. The approach in these papers to weighting for case difficulty has been implemented in California and Washington. Most state agencies experimenting with weighting appear to be using variations of the difficulty indices developed by Duane Sermon. Cf. Duane T. Sermon, The Difficulty Index - An Expanded Measure of Counselor Performance, State of Minnesota Department of Vocational Rehabilitation, Research Monograph No. 1, March, 1972.

A distinction should be kept in mind here, however, between "weighting" case closures with expected success or cost predictions so that the effectiveness of services can be evaluated, and "weighting" case closures to give more credit for serving more difficult cases so that counselors will shift their caseloads in the direction of more severely disabled individuals. The former type of "weighting" changes evaluation such that "creaming" does not result in extra credits and thus evaluation becomes neutral to the case-mix choices of the counselor. The latter kind of "weighting" intentionally attempts to influence the case-mix choices of counselors in particular directions and may even result in undermining any evaluation of efficiency or effectiveness in serving clients. For example, if counselors are given more credit the higher the cost of the rehabilitation or the longer the time in process of the client (in the name of increasing services to the severely disabled), there is no incentive to select the most effective or efficient service-mixes or strategies for a client having given set characteristics and needs.

8. Follow-up Data. If an evaluation unit has any capacity to move beyond routine processing of R-300 data for its evaluation efforts, it should put first priority upon the collection of follow-up data for sample of those client groups with which it is concerned to demonstrate good performance. Follow-up data on the client's situation at some point in time following closure is critical if the validity of R-300 indicators as program success measures is to be accepted, both by outsiders and by insiders. Once obtained, follow-up

data can also be used to determine which indicators should receive priority in the future and which indicators need supplementation by non R-300 data. For example, it will normally prove true that the higher the wages received at closure, the higher the probability that the 26 closure will continue to be employed and to have experienced wage increases in a post-closure survey. It may also prove true that those 26 closures with lower wages at closures who retain employment or experience advancement can be predicted by their occupational placements or by other characteristics (e.g., education, age). Once follow-up data is compared with closure information and client characteristics, it is possible to derive procedures for using R-300 closure information with more confidence in future evaluations of program impact.

A few suggestions for follow-up may be useful in reducing the costs of such studies. For most purposes, it appears that a follow-up of the client's situation a year after closure will be sufficient to pick up the exodus of most 26 closures from the labor force. Review of three- and five-year surveys (which are few) suggest that the percentage of 26 closures still employed stabilizes after the first year. In a follow-up survey, it is insightful but not absolutely necessary to monitor the client's employment experience for the whole time following closure. Many of the flows into and out of the labor force cancel out, especially when homemaker closures comprise a fair proportion of the rehabilitants. Besides, the interviewee's recall of

his past employment is often highly problematic, especially in situations where jobs change several times. If costs must be minimized and the length of the interview or questionnaire reduced, it is most important simply to know the current employment status and weekly wages of the interviewee, whether he believes that changes in his employment status following closure are in any way attributable to the training and help he received from the rehabilitation program, and perhaps how long following closure the client worked in an occupational field corresponding to the training received or to the type of job held at closure. Finally, the costs of surveys can be kept down by using phone or mail surveys; mail surveys should not be used, however, unless coupled with telephone follow-up to reduce the nonresponse rate.

9. Cost data and Specification of Program Activities. In evaluation, especially in pseudo-applications of the experimental design model, the nature of the program's intervention is too often taken for granted. The "program" is treated as a "black box" with little attention to what actually takes place with the client. Changes in the before-after situations of clients and perhaps some reference group are measured and used to impute the success of the "program" (usually an office or a counselor or an agency) which reputedly served the clients. Unfortunately, the rehabilitation model emphasizes a highly individualized counselor-client relationship with the specification of an integrated mix of services carefully tailored to the particular needs of each client. This model is the norm, of

course, not the reality in many cases. Yet the norm and reality is such that the delivery of rehabilitation services is not readily treated as a standardized process. Evaluations of offices or program strategies using before-after data on clients should be coupled with process analysis, case studies, and even some file and account audits to see what actually took place during service delivery. Such eclectic ad hoc analytic approaches can often provide the insights about the reasons behind poor or good program performance, so that decision-makers can make the right decisions about what aspects of the program should be copied elsewhere or reformed.

Quite apart from knowing what happens during service delivery, it is also important to estimate how much services really cost and whether services which were successful were being supplemented by services being delivered by other agencies. Unfortunately, in rehabilitation, we currently only monitor a portion of the costs which vary with the client - case service expenditures. Counselor's time, the fixed costs in plant, training, and administration, and the case service expenditures incurred in serving clients whose cases were closed unsuccessfully are allocated across the entire caseload of the agency on some arbitrary basis. Many states do not even adjust total fiscal year cost data in estimating the costs of that year's 26 closures, to reflect the changing size of the "still active" caseload. As a consequence of these practices, there is little information on the true program costs of serving different kinds of clients. At the same time, case service costs per rehab are often used to assess

performance with different disability groups even though such costs constitute only about 25% of total costs and probably 50% or so of true variable costs. These kinds of problems are readily familiar to any cost analyst, but administrators who have come up through professional counseling backgrounds are often unfamiliar with management and cost accounting and why accurate cost information tied to individual agency "products" (in this case, rehabilitations) are important. An effective evaluation unit must be sensitive to the implications of alternative assumptions about the distribution of costs. Periodic monitoring of how samples of counselors spend their time among activities and among different clients could also be useful in developing formulae for allocating the very large block of costs represented by counselors' salaries. The linkage of R-300 data to the agency's accounting data will also be useful in assessing the effectiveness of different kinds of case service expenditures (e.g., college training vs. workshop training vs. restoration vs. diagnosis) and of greater or less expenditure on a particular kind of service, controlling of course, for the client's characteristics.

It is not enough to focus only on costs borne by the agency in assessing program outcomes, however. The costs of services attributing to the outcome of the rehabilitation process are in some cases shared to a varying extent by other agencies and by the client and his family. If another agency is providing supplementary services in a given district to rehabilitation clients, it is silly to believe that doubling the size of the rehabilitation offices in that district will double the number of closures produced at the same cost per

closure, unless there was some reason to believe that the other agency was also capable or willing to double its services.¹ Yet rehabilitation agencies are often surprisingly indifferent to the activities of other agencies, and vice-versa. (In a world of indifference, of course, both agencies with a clear conscience can simultaneously count the client's improved situation as uniquely the consequence of its own services.) The share of costs borne by the client and his family (direct expenses for rehabilitation services, expenses on auxillary services - like transportation and child care - that allow the client to receive rehabilitation services, and any employment earnings that may be foregone while the client is in the rehabilitation process) may also be significant and may explain in some cases why individuals drop out of the program or do not succeed.²

¹ There also is no reason to assume without further information that enough clients with disability characteristics of similar difficulty to those currently being served will exist within the district to allow the program's capacity to double with no change in program performance. Presumably, if "creaming" is widespread, then increases in program scope in any given area which are not marginal, should result in the program coming into contact with relatively more severe disability and thus more difficult cases. Similarly, there is no reason to assume that the area is capable of attracting enough counselors and supervisors of comparable skills and effectiveness to those currently employed. The problems here are the classical ones of failing to distinguish between the measured success of the program in dealing with the "average" or typical client in the caseload and the success in dealing with the "marginal" client being added to the caseload as the program expands.

² It is also useful to view benefits from the client's perspective. Where the probable job awaiting after services offers only a low wage with limited career opportunities, some clients on welfare may find little financial gain in substituting the job earnings represented by "rehabilitation" for welfare payments that can be received without

10. Focusing Evaluation upon Decisions or Alternatives. The evaluation unit which is likely to convince quickly the state director that evaluation is something deserving agency commitment is the evaluation unit that focuses evaluation upon providing information necessary for particular decisions or for assessing alternative strategies or actions for improving performance. Evaluations that simply document the performance of the overall program are primarily useful for program defense, not internal management or improving performance. Evaluations which document the performance of the various district offices or counselors within the state may be useful in helping line administrators decide where the trouble-spots are, but they do not provide specific insight concerning the source of trouble.

Evaluation units would be well advised continually to formulate the questions they seek to analyze in terms of input-output relationships, that is, to ask what are the outputs or outcomes associated with alternative inputs into the rehabilitation process. Inputs consist of case-mix selection, service-mix selection, the amount of expenditure on particular services or in toto, the type of counseling, the training and experience of the counselor (as a proxy for the "quality of counseling"?), the type and amount of administrative support and supervision given the counselor, the availability, quality, and exploitation of supplementary services by other public and private agencies,

engaging in dead-end work. This suggests that in providing vocational training for welfare recipients, it may be necessary to train individuals for jobs paying much better than minimum wage if incentive for rehabilitation is to be provided.

the cooperation and active participation in the rehabilitation process of the client's family, community attitudes, the occupational structure and overall level of the local economy, public assistance practices in the local community, and so forth. Obviously, it is most useful to function on inputs which the rehabilitation program has some ability to manipulate in the short run and thus to use if appropriate to increase performance in achieving goals. Still, it is also helpful to keep in mind those factors which are beyond the control of the program but which perhaps condition whether or not agency-manipulated inputs will be effective.

If an evaluation unit does learn to formulate problems in such input-output terms, they will more often produce information when analyzing problems or decisions, which may be useful in improving performance. They also will help generate an environment inside the agency which is conducive to experimenting and learning, rather than attempting to conceal or assign "blame" for failures. There is no need to pretend that someone really does have knowledge of how to rehabilitate or solve every client's employment and living problems, and thus to insist that sub-par program performance indicates professional incompetence or laziness. The image of the all-knowing professional rehabilitation counselor has long deserved to be abandoned. We have no reason to be ashamed of the fact that we sometimes fail (and fail mightily). We do have reason to be ashamed as "professionals" if we are unwilling or unable to recognize or admit when we have failed, and look for better ways of serving clients. The issue

investigated in evaluation should not primarily be whether the program is performing well or poorly, but rather whether and how performance can be improved.

11. Use of Reference Groups. Knowing the level of a program's or a program subcomponent's performance is not very useful unless there is some standard or reference group with which the level can be compared and thereby assessed. Ideally, there would be a standard by which evaluators would know whether a given performance was "good" or "bad", or - more to the point - "good enough" or "not good enough" to conclude that corrective action is not necessary. Where objectives and targets, reflecting goals, are clearly set, there is a standard for assessing performance. If the target was X rehabilitations of a certain disability during the fiscal year, did the number of rehabilitations exceed or fall beneath the target? Even with such specific targets, however, there is still the nagging question of how can we tell whether the targets we set were themselves too high or too low, given our capabilities. Clearly, setting standards in terms of rates which are independent of size (e.g., percentage of need met, percentage of cases accepted which are rehabilitated, cost per rehab, rate of return on investment) is more useful than setting standards in terms of absolute numbers. An agency with twice the budget of another agency should have more demanded of it. The use of rate-standards adjusts for scale. The problem still remains to set the standard, however.

In practice, we tend to forego the search for absolute standards and to evaluate instead by comparing program performance to that of reference groups. If we perform better than the reference group, we consider that we are doing well. As noted before, in making such comparisons, it is necessary to control for scale and expenditures. Among the reference groups available for comparison with state rehabilitation agencies in terms of performance are the following:

a.) Internally generated, randomly assigned groups of clients who were found eligible but were not served. This is the ideal, of course, of the experimental design model for evaluation. Such "control groups" might be feasibly generated by putting eligible applicants on wait lists and simply delaying the start of services. Since the rehabilitation program has far more applicants than it can serve, such a systematic use of a wait list for experimentation need not be viewed as an unfair denial of service. It should be noted that to the extent that the control group truly receives no services from any source, a comparison of outcomes informs us whether services have had any impact. We do not learn particularly whether the services as rendered in this particular program were more or less successful than similar services rendered elsewhere. We only can learn whether rehabilitation services are better than no services at all.

b.) Status 30 Closures. Here one compares follow-up data on 30 closures to follow-up or closure data for 26 closures. At first glance, this reference group looks similar to a true control group.

It consists of individuals who have been declared eligible (and thus comparable to the clients served) but who drop out before significant amounts of service are received. Unfortunately, such individuals also tend to have less severe disabilities, higher motivation, and characteristics more associated with labor market success than the characteristics of 26 closures.¹ The group also may receive substitute services from other sources. Many of these individuals, however, drop out because they have found jobs on their own. Using them as a comparison group without further adjustments tends to understate the actual achievements of rehabilitation services.

c.) Status 28 Closures. Here presumably one is interested in comparing follow-up data for 26 and 28 closures, program successes and failures. One expects to find obviously that successful closure should have better employment situations than unsuccessful closures. Again the evaluator must be careful to adjust the reference group to make it comparable. Status-28 closures generally tend to be more severely disabled and handicapped (at referral), to have characteristics less frequently associated with labor market success, and perhaps to have less motivation and more personality problems (at referral) than 26 closures. Using them as a comparison group without further adjustments tends to overstate the actual achievements of rehabilitation services. Again, it may also be true that individuals closed as 28's

¹ Thomas K. Glennan. Evaluating Federal Manpower Programs: Notes and Observations. Rand Corporation, Memorandum PH-5743-0E0, (September, 1969).

will subsequently receive post-closure services from other services which may improve their situation. Also, some of the post-closure gains experienced by 28 closures may, by their own admission, be attributable to the rehabilitation services they earlier received.

d.) Status 08 Closures. Here one compares follow-up data for 08 closures with closure or follow-up data for 26 closures. Again at first glance, it appears that 08 closures provide a group similar to the experimental control group, a group which has not received services. The similarity is very deceptive, however. The 08 closures were in fact rejected for services either because (1) they were too severely disabled to be judged feasible for vocational rehabilitation, or (2) they had no disability severe enough to make them eligible for services. In the case of the first subset of 08 closures, like the case of 28 closures but even more so, use of the subset as a reference group tends to overstate the actual achievement of rehabilitation services. In the case of the second subset, like the case of 30 closures but even more so, use of the subset as a reference group tends to understate the actual achievement of rehabilitation services. Combining the two subsets into a single reference group makes comparison with 26 closures even more absurd. Again, the problem also exists and indeed is even more likely than with 30 or 28 closures, that the 08 closures will have received services from some other source which influenced their employability.

e.) Other Rehabilitation Programs Operating in the Same Time Period. One of the best frequently used reference groups by state

legislatures for evaluating the performance of their own state program is the collection of other state programs, as represented by the average performance across all states or by the rankings of all states in terms of published R-300 indicators. Although the state director is quick to note that his program operates in a different economy with a different type of caseload and under different operating and administrative conditions, the continued availability of published Federal data giving such rankings encourages legislators to use such standards for evaluation. Similarly, state agencies frequently evaluate a given district's or counselor's performance by comparing it with the performance of other districts or counselors within the state. As noted in our discussion of indicators, if the indicators were first adjusted for case-mix, their use for evaluation of either states or district offices could be appropriate. Such use is more appropriate with districts and counselors within state offices than with state agencies. Within states, program subcomponents operate under similar administrative and environmental conditions. The differences in operating conditions between urban and rural areas within a large state thus are presumably less than the differences in operating conditions between urban and rural states;

f.) Variations among clients within the rehabilitation program in terms of program contact. One of the most overlooked reference groups for a rehabilitation program are its own clients. Clients in rehabilitation programs, even within subprograms, vary greatly in the type and amount of services received. Similarly, there are usually

many natural experiments within any program where fairly different strategies of service delivery are practiced. Variations in outcome can be compared among clients with variations in services received or strategies of service delivery experienced using existing agency files. Limitation of this approach, however, is that to the extent that all clients share the experience of certain kinds or a minimum level of service (e.g., some counseling, diagnosis, simple contact with the program), not all of the variance in client outcome can be allocated among the kinds of services or delivery strategies experienced. Moreover, the services clients receive and the program sub-components in which they participate are not random. The client's needs (as proxied by his characteristic background, and disability) determine the services received. Thus, evaluations based on variations in client outcomes and services received within the program must again control for client characteristics.

g.) Past historical experience. Another frequently used standard for assessing program performance is past program performance. Change in indicators in the program are analyzed across time. Again, it becomes important to adjust for any changes in caseload, major operating conditions, and record-keeping practices and definitions which occur across time. Many of the "rules of thumb" and the "good professional practice" procedures that are widely used to evaluate programs and counselors have their basis primarily in his-

torical experience and tradition.¹ Whether historical experience and tradition were indeed "good enough" to provide a standard by which current performance should be judged is seldom explicitly considered.

h.) Other Programs pursuing goals similar to those of rehabilitation programs. Occasionally, analysts in state Finance Departments or legislatures will seek to compare the performance of vocational rehabilitation programs with the performance of other related programs (e.g., manpower programs, public offender rehabilitation programs). Such comparisons are even more problematic than use of the other reference groups cited, because of the major variations that usually exist in the caseloads, administrative and operating conditions, and data systems of different program systems.

Clearly, there are different problems and biases associated

¹ Initially the "good professional practice" procedures were probably based on theories about human and social behavior. The theories themselves were seldom formally tested, much less verified. At best, it might be argued that the procedures became widely accepted as standards within their own time because they appeared "to work" and thus pragmatically were used. A reading of the literature of many service fields (e.g., social work, education, rehabilitation counseling) suggests, however, that the procedures often were not perceived as having "worked" by practitioners. The source of the thrust for evaluative research in many of the service fields has been precisely the recognition that more formal experimentation was required to determine what practices are effective. Thus, the acceptance of these untested "good professional practice" standards, as reflected for example in the use of manual guidelines to develop standards for audits of counselors' case records and case work, appears to be based more on the fact that the practices have long been followed rather than upon their theoretical soundness or validity. We thus shall classify such standards as using the historical program as a reference group.

with using different reference groups to provide a standard for judging program performance. An effective evaluation unit must be sensitive to these problems and be careful to make the adjustments necessary to insure comparability.

IV. Evaluation Tools - Methods and Methodological Approaches

Quite apart from the above organizational and functional requirements that must be met by any evaluation unit that seeks to use evaluation effectively to improve organizational performance in achieving goals, the unit must also have at its disposal knowledge and skills in study design, sampling, survey research, data processing, and data analysis (not just statistics). Books have been individually written for each of these tools and it makes no sense to pretend here that cookbooks are simply made available. Rather than futilely attempt to synthesize in a few pages what evaluation units need to know about sampling or analysis, I have chosen to emphasize instead the acquisition of staff with training and experience in such tools. These methods and skills are generic to many programs and many organizational functions (e.g., planning, research and development), rather than peculiar to the needs of rehabilitation programs or evaluation. No evaluation unit and no evaluator will ever be sufficiently knowledgeable in all these areas, just as no university or consulting team is ever fully knowledgeable. Indeed, in practice, evaluators will seldom be able to apply rigorously the methodological knowledge that they do have. Evaluation research is necessarily full

of problematic assumptions, imperfect reference groups for comparisons, peculiar combinations of hard and soft data analysis, and the exercise of hopefully good judgment in areas where data is missing or rigorous research design is infeasible. Analytical training and experience, however, sharpens the evaluator's judgment and enhances the quality of his eclectic choice of method.

Quite apart from specific methods and research skills, there are many different kinds of methodological approaches which can be used in evaluation. While some of the literature would appear to equate program evaluation with evaluative research and experimental design, anyone observing how evaluation is actually carried on will observe limited use of experimental design and wide use of site visits, case studies, testimonials from clients and other agencies, and file audits as methodological approaches to evaluation. These latter approaches can be dismissed as "non-scientific" but such a dismissal is not very useful. What we need to know are the strengths and limitations of each of these approaches in discovering evaluative information which can help agencies make the right decisions or improve performance. Similarly, several rehabilitation agencies and many non-rehabilitation programs have made effective use in evaluation of other seemingly "sophisticated" methodological approaches which have at best an indirect relationship to experimental design:

benefit-cost analysis

operations research, simulation, and systems analysis

input-output analysis using quasi-production functions

process analysis
process consultation
consumer evaluation
peer review
goal attainment scaling
... and so on.

Each of these approaches tends to be useful for shedding light on some kinds of questions or problems and to be less useful for other problems. These methodological approaches represent evaluation tools just as do sampling, particular quantitative methods, and survey design.

This paper has thus only begun to scratch the surface in outlining "tools" available to evaluators. Evaluation is a young field with much development yet to take place. The comprehensive how-to-do-it manual for program evaluation is not yet and probably never will be feasible. Certain basic requirements for program evaluation do exist, however, and these we have tried to describe.