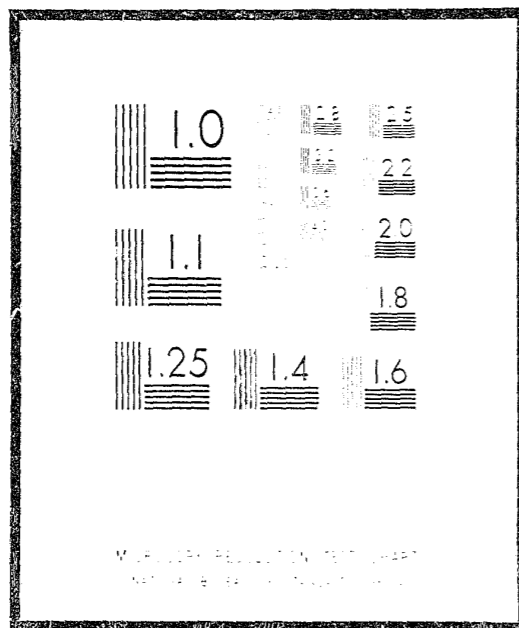


NCJRS

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



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

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

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

Date filmed

5/27/76

O 01 401 TEL 18 50

PROGRAM EVALUATION - EMERGING ISSUES
POSSIBLE LEGISLATIVE CONCERN RELATIVE TO
THE CONDUCT AND USE OF EVALUATION IN THE
CONGRESS AND THE EXECUTIVE BRANCH

CONGRESSIONAL RESEARCH SERVICE LIBRARY OF CONGRESS

19846

EXECUTIVE DIRECTOR OF
ANALYSIS, SCIENCE, AND TECHNOLOGY
POLICY RESEARCH DIVISION
November 16, 1976

TABLE OF CONTENTS

	Page
ABSTRACT	
PART I. INTRODUCTION.....	1
PART II. THE INITIATION AND USE OF PROGRAM EVALUATION.....	7
A. The Need to Clarify Legislative Statements of Program Goals and the Purposes of Evaluation.....	9
1. Discrepancies between the Purposes of Evaluation and the Techniques of Evaluation.....	9
2. Recommendations to Specify Program Goals in Legislation.....	11
3. Criticisms of Specifying Program Goals in Legislation.....	13
B. Better Communication for Effective Evaluation.....	14
C. The Integration of Program Evaluation into Program Management.....	17
1. Incorporation of Evaluation into Agency Program Planning and Budgeting.....	17
2. The Development of Evaluation Work Plans.....	19
3. LEAA: An Illustration of These Deficiencies.....	19
4. The Contribution of Evaluation Work Plans to Evaluations Performed at the State Level.....	20
5. Proposals for Comprehensive Evaluation Plans.....	21
6. Proposals for a "Preassessment of the Evaluability of Programs".....	22
D. The Need for Better Dissemination of Information on Program Evaluations Supported Under Federal Agency Auspices	23
1. Investigation by the General Accounting Office.....	24
2. NIMH/UCLA Project on Data Bank of Program Evaluations (DOPE).....	24
PART III. IMPROVED COORDINATION OF PROGRAM EVALUATION IN THE EXECUTIVE AND LEGISLATIVE BRANCHES.....	26
A. Issues Surrounding the Need to Incorporate Program Evaluation Findings into Executive Branch Budget Processes....	26
B. Proposals to Give the Office of Management and Budget Program Evaluation Responsibilities.....	29
C. The Proposal to Establish a Federal Program Evaluation Office in Either the Office of Management and Budget or in the Executive Office of the President.....	29
D. Current OMB Initiatives in Program Evaluation.....	30
E. Issues Relating to the Improvement of Legislative Branch Program Evaluation Capabilities.....	32
F. Issues Surrounding the Role of the General Accounting Office in Providing the Congress with Improved Program Evaluation Capability.....	36
1. Basis of GAO's Statutory Authority for Program Evaluation.....	36
2. Evolution of GAO's Program Evaluation Activities.....	36

TABLE OF CONTENTS (cont.)

	Page
3. Illustrations Provided by the GAO of Its Program Activities.....	38
4. Other GAO Program Evaluation Activities.....	39
5. Activities and Legislation in Support of Giving GAO More Responsibility for Program Evaluation.....	39
PART IV. THE PROCUREMENT AND CONDUCT OF PROGRAM EVALUATION RESEARCH...	41
A. Issues Relating to the Funding of Program Evaluation Research.....	42
1. The Lack of Appropriate Information.....	42
B. Difficulties of Obtaining Information on Types of Performers of Program Evaluation Research.....	45
C. Specific Issues Relating to Federally Funded Research and Development Centers as Performers of Program Evaluation Research.....	47
D. Issues Relating to Funding Mechanisms (Grants and Contracts) in Program Evaluation Research.....	49
1. Issues Relating to Profit- Versus Non-Profit-Making Performers.....	49
E. Legislative Requirements to Mandate a Certain Percentage of Funding for Program Evaluation.....	50
F. Alternative Institutions for the Conduct of Program Evaluation Research Under Federal Agency Auspices... ..	51
PART V. METHODOLOGICAL ISSUES IN THE CONDUCT OF PROGRAM EVALUATION....	58
A. Experimental and Non-experimental Research Designs.....	58
B. Scope and Limits of Social Experimentation.....	59
1. Legislative Interest in the Ethics of Experimentation.....	65
C. The Need for Continuous, Appropriate Impact Data.....	65
D. Cross-fertilization Between Advances in Social Indicators Research and Improvements of Data Required for Program Evaluation.....	69
PART VI. THE POSSIBLE ISSUES OF LEGISLATIVE CONCERN RAISED BY INFORMATION PRESENTED IN THE REPORT.....	75

ABSTRACT

The conduct of program evaluation research has grown considerably in recent years, in response to explicit legislative mandate as well as executive agency directive. However criticism abounds about the utility to the government of many of these types of research. This report discusses some of the factors --- both governmental and non-governmental which contribute to this situation. It also identifies recent actions taken to remedy the discrepancy between the promises and utility of evaluation research. Remaining issues of possible legislative concern are also listed.

PART I. INTRODUCTION

Federal expenditures for program evaluation are estimated to have risen by more than 500 percent from 1969 to 1974, from \$20 million to more than \$130 million. This increase in program evaluation activities reflects two important trends. The first is a shift in total Federal budget expenditures toward greater emphasis on human resources programs, from approximately \$65 billion in 1969 to almost \$173 billion in 1976. 1/ The second trend consists of a "many-fold multiplication" in congressional appropriations for evaluations in response to increasing demands from the Congress, the President, and executive agencies for greater accountability and evaluation of the resources devoted to social programs. 2/ The conduct of program evaluation is also part of a broader trend to provide decisionmakers in both executive and legislative branches of Government with objective information or policy advice on the effects of Federal actions and programs, preferably before large-scale application. 3/ Dr.

1/ Executive Office of the President. The United States Budget in Brief, Fiscal Year 1976, Washington, D.C., U.S. Govt. Print. Off., 1975. p. 7.

2/ Mushkin, Selma J. Evaluation: Use With Caution. Evaluation, v. 1, 1973: 31. The Senate Agriculture Committee's needs for evaluation especially an evaluation of the Farmer's Home Administration are described in James W. Giltmier. Policy Formation Through Program Evaluation and Systems Analysis: A Congressional View, mimeo, 20 pages, paper presented before Evaluation Seminar, (held monthly under Sen. Brock's authority), Seminar held November 20, 1974, U.S. Capitol.

3/ Program evaluation activities required under the Congressional Budget and Impoundment Control Act are described below in this Introduction. In addition, The Office of Technology Assessment, established under P.L. 92-484, October 13, 1972, as an information and analysis office attached to the Congress, will "...provide [the Congress with] early indications of the probable beneficial and adverse impacts of the applications of technology..." and other relevant information, to assist in the legislative process. [See: U.S. Congress. Senate. Committee on Rules and Administration. Subcommittee on Computer Services. Technology Assessment for the Congress, staff study, November 1, 1972, 92nd Congress, 2nd session, Washington, D.C. U.S. Government Printing Office, 1972. (Committee Print.)] Judicial interpretation of the National Environmental Policy Act of 1969, now reflected in revised guidelines issued by the Council on Environmental Quality, requires the submission of 102 impact statements on a wide variety of major Federal activities which may impact on the "human environment," broadly defined. These activities include construction of facilities, alterations to the natural physical environment, funding proposals, and administrative actions, as well as proposals for different types of Federally supported scientific research and

H. Guyford Stever, the President's Science Adviser estimates that \$835 million will be obligated for social R and D programs in the fiscal year 1975. In fact in discussing the 1975 budget message he singled out for special consideration as social experiments designed to generate information for public policymaking, three programs in health insurance, income maintenance, and housing allowances, with obligations estimated to total \$33.9 million. ^{4/}

Concurrent with increasing activity in social program evaluation, however, there is evidence of heightened concern by Members of Congress, agency program administrators, and social scientists, with inadequacies in program evaluation methodology, procedures for data collection, procurement, administration, and utility of program evaluation, especially in planning, oversight, and budgeting functions.

The history of Federal "social program evaluation" activities has never been fully chronicled. Similarly, there is no detailed analysis of the merits of evaluation objectives, programs, and approaches used by different social program-oriented agencies. This report is designed to provide the Congress with background information on existing and emerging issues of concern relating to evaluation. Part I is this introduction. Part II deals with some general issues in program evaluation: its objectives, communications between researchers and program managers, and the difficulties of using evaluation in agency management procedures. Part III treats OMB program evaluation activities, proposals for interagency evaluation mechanisms, and activities surrounding improvement of legislative branch programs evaluation capabilities. Part IV deals with issues relating to the procurement and conduct of program evaluation research from the point of view of both the evaluation researcher and the government user. Part V deals with methodological issues in the conduct of program evaluation. The possible issues

development which may result in development of technology with a potential impact on the environment. (See: Council on Environmental Quality. Preparation of Environmental Impact Statements, Guidelines. Federal Register, v. 38, no. 14, Part II, August 1, 1973).

^{4/} The Science Adviser, "Federal Budget for Science and Technology - FY 1975," February 4, 1974: 8.

of legislative concern raised by information presented in the report are collated in Part VI.

The term "program evaluation" is used in two general ways in the Federal government. At a very aggregated level, the term denotes a process of assessing cost and effectiveness to ascertain the appropriate funding level for a program in relation to other national goals. At a less aggregated level, the term refers to the processes of evaluating a particular agency's program to determine whether and to what extent it achieves its purposes (impact evaluation), and whether it is administered efficiently (process evaluation). Decisionmakers use these evaluations in deciding whether a program should be continued, modified, terminated, or replaced by another "more beneficial and more effective" program.

The term "program evaluation" has been associated with any activity to measure the effectiveness and outcome of a program. However, it has been most widely associated in recent years with evaluation of major social programs, such as educational intervention strategies, work incentives, and juvenile training. In this respect, evaluation, according to Wholey, et. al., provides "objective information to program managers and policy-makers on the costs and effects of national [social] programs and local projects, thereby assisting in [promoting] effective management and efficient allocation of limited resources." ^{5/}

Federal agencies have developed a number of different types of program evaluation mechanisms. The Congress mandates some evaluations, which are conducted at stipulated funding levels and with mandated techniques. Some agencies initiate evaluation to provide information for program planning and budget justifications. Effective coordinated program evaluation offices exist in some agencies. Other agencies have not organized their program evaluation activities formally.

^{5/} Wholey, Joseph S., et. al. Federal Evaluation Policy: Analyzing the Effects of Public Programs. Washington, D.C., The Urban Institute [1970] pp. 19-20.

One characteristic is common to almost all social program evaluations --- that is methodology. Thus according to Wholey, et. al., "evaluation is research, the application of the scientific method ... to learn what happens as a result of program activities." Evaluation thus includes:

the definition of program objectives, the development of measures of progress toward these objectives, the assessment of what difference public programs actually make, and the projection of what reasonably could be expected if the program were continued or expanded. 6/

The tools and techniques of program evaluation usually incorporate the basic techniques of social science research --- systematic collection of evidence from a representative sample of the population; translation of information into quantitative terms; and analysis of the information, using statistical manipulation and reference to theoretical notions about human behavior, to "draw conclusions about the effectiveness, the merit, the success, of the phenomenon under study." 7/

Much of the current criticisms of program evaluation research relate to its state-of-the-art and its questionable utility to the government. A number of recent or on-going governmental and professional activities illustrate these concerns.

The need to improve the state-of-the-art of program evaluation is underscored by the establishment within the Research Applied to National Needs program (RANN), of the National Science Foundation, of a \$2.5 million program to provide "rigorous examination" and refinement of the techniques of program evaluation and policy-oriented research in the fields of labor, youth, health, and criminal justice. 8/

The Science and Technology Policy Office of the National Science Foundation is supporting three research projects on social science policy, designed in part to improve the conduct and use of program evaluation. These studies are:

6/ Idem.

7/ Weiss, Carol H. Evaluation Research: Methods for Assessing Program Effectiveness. Englewood Cliffs, Prentice-Hall [1972], p. 2.

8/ "NSF Initiates Programs for Evaluation of Policy-Related Research Concerned with Human Resources." NSF News Release, April 11, 1973. (NSF 73-137.)

National Academy of Sciences, Study of Social Research and Development and its Role in Policy Making, 6 mos., \$116,500;

Social Science Research Council, Conference on Social Experimentation, 12 mos., \$23,100; and

University of Washington, Social Policy Analysis and Research in Social Policy-making, 14 mos., \$50,000.

Some of these topics were addressed during two conferences held at Dartmouth during the summer of 1974. The first, a conference on Social Experimentation, was sponsored by the Social Science Research Council with funding from the National Science Foundation. The second, the "Dartmouth/Organization for Economic Co-Operation and Development (OECD), Seminar on Social Research and Public Policies," was supported with funds from the Ford Foundation, Dartmouth, and the OECD. The Fifth Annual Conference of the National Capital Area Chapter of the American Society for Public Administration had a session devoted to examining the conduct and utility of program evaluation. It was held on November 14 and 15, 1974 in Washington, D.C.

Numerous bills were introduced during the 93rd Congress to incorporate rigorous program evaluation, pilot test, and cost-benefit analysis procedures into legislation in order to improve congressional oversight of the budget. The Congressional Budget and Impoundment Control Act of 1974 (P.L. 93-344, approved July 12, 1974), established a Program Evaluation Office in the General Accounting Office and gives the agency added responsibility for program evaluation. It also strengthens legislative responsibility for using program evaluation information.

The topic of assessing the quality of evaluation research, especially experimentation, for decisionmaking processes is also receiving some attention by the Special Committee on Federal Agency Evaluation Research of the Division of Behavioral Sciences of the National Research Council. The Office of Management and Budget has also established an Office of Evaluation and Program Implementation to assess these issues. In addition, the General Accounting Office has published several reports directed to congressional committee chairmen, recommending steps to be taken to refine legislative language

requiring program evaluation. The GAO is now preparing a guide for program evaluation standards, expected to be released in the Spring of 1975.

An effort is made in this report to present background information in a manner which serves the interests of the Congress. Many of the issues raised may be relevant to both House and Senate Budget Committees and the Congressional Budget Office, established under the Congressional Budget and Impoundment Control Act of 1974, P.L. 93-344 of July 12, 1974. In addition some authorizing committees as well as committees responsible more generally for overseeing the organization and administration of Government might possibly be concerned with the issues raised.

The following format is used in presenting the material. Each issue is developed by referring to current activities and literature, such as those of the Office of Management and Budget, the General Accounting Office, the National Science Foundation, and the National Academy of Sciences; social science critiques; and published and unpublished Federal agency reports. Also drawn upon if applicable, are legislative activities which address these topics. If available, recommendations for improvement from published sources are given. The possible issues of legislative concern generated by the report are presented in Part VI.

(It should be noted that the Library Services Division of the Congressional Research Service has published multilithed annotated bibliographies on program evaluation. All were prepared by Nancy Davenport. They are: Evaluation Research in Public Administration: Selected References, 1967-1974, October 21, 1974, 9 p.; Evaluations of Manpower Training Programs: Selected References, 1970-1974, October 25, 1974, 8 p.; and Evaluation Research in Social Policy: Selected References, 1970-1974, October 31, 1974, 8 p.)

PART II. THE INITIATION AND USE OF PROGRAM EVALUATION

The General Accounting Office (GAO), estimates that "Federal expenditures for non-defense program evaluation have risen dramatically in the past few years --- from less than \$20 million in the fiscal year 1969 to at least \$110 million in fiscal year 1972 ^{9/} and \$130 million in the fiscal year 1974. ^{10/} As a result, according to the GAO:

In response [to the increasing demand for program evaluation] executive agencies have set up new evaluation offices and expanded existing offices. Contracting for evaluation studies has expanded greatly. The trend in number and size of evaluation studies sponsored by the Federal Government has been upward. Private foundations as well have stepped up funding of evaluation studies. State and local governments have increased evaluation efforts and several universities have set up evaluation institutions. ^{11/}

Although program evaluation is extensively used, there is widespread criticism of its utility to decision-makers. For instance, Sen. William V. Roth, Jr., reporting to the Congress on a 1972 staff study entitled Report on Survey of Federal Program Evaluation Practices, said "the study conducted by my staff suggests serious weaknesses in agency evaluative and analytical procedures." The report, he continued, "call[s] attention to the need for the executive branch to improve and extend its attempts to

^{9/} U.S. General Accounting Office. Program Evaluation: Legislative Language and a User's Guide to Selected Sources. Washington, D.C., U.S. General Accounting Office [June 1973] p. 1. It should be noted that most of the literature on program evaluation indicates increasing demand by both Federal agencies and the Congress, and increased funding by agencies to support the conduct of program evaluations. Dr. Allen Schick disagrees. He reports in a 1973 publication: "...There has been a definite drop of interest in large program evaluation, though many Federal agencies are trying to institutionalize evaluation through the collection of output and performance data." (Allen Schick (Senior Specialist, Congressional Research Service, Library of Congress), "The Pilot Testing of New Programs: An Analysis of Title IV, S. 3984 (92nd Congress)," In U.S. Congress. Joint Economic Committee. Subcommittee on Budgeting, Management, and Expenditures. Improving Congressional Control Over the Budget: A Compendium of Materials, March 27, 1973. Washington, D.C., U.S. Govt. Print. Off., 1973. p. 276. (Committee Print.)

^{10/} "Program Evaluation in the Executive Branch: Estimate of Funds Budgeted for Fiscal 1974," in U.S. Congress. Joint Committee on Congressional Operations. Congressional Research Support and Information Services: A Compendium of Materials. Washington, U.S. Govt. Print. Off. [May 1974] pp. 289-290.

^{11/} Program Evaluation: Legislative Language and a User's Guide to Selected Sources, op. cit.: p. 1.

measure the accomplishments of governmental activities and weigh these accomplishments against their costs." ^{12/} Federal administrators who fund and oversee program evaluation research have also criticized its utility. For instance, Robert G. Bruce, Assistant Administrator for Program Planning and Evaluation, Social and Rehabilitation Service, HEW, reports: "We might as well be candid: Federal program evaluations so far have been largely ineffective." ^{13/} Garth Buchanan and Joseph Wholey, program evaluation researchers at the Urban Institute, a research and evaluation agency created first to serve the Department of Housing and Urban Development, observe similarly:

... While we have seen that increasing amounts of money are being budgeted for evaluation, we conclude that this increasing support must be due to a continuing recognition by the Government that evaluation information is needed for the development and management of social programs, rather than to the recognition that evaluation as currently practiced is the answer to these needs. We are led to this conclusion because in our judgment, the impact of evaluation results on program development and improvement over the last two years has been disappointing when compared with the amount of money and effort that has gone into evaluation. ^{14/}

The Office of Management and Budget recently initiated a major effort to improve Federal agency program evaluation capabilities. Agency spokesmen identify four basic difficulties with current Federal program evaluation research. Their concerns parallel those raised above. They are: (1) timeliness for decisionmaking; (2) reliability of the data used and the methods employed; (3) relevance of evaluation activities to major policymaking needs; and (4) insuring that program evaluation reports are made available to and utilized by decisionmakers. ^{15/}

^{12/} Public Program Analysis and Evaluation for the Purposes of the Executive and the Congress. Statement of the Hon. William V. Roth on the Floor of the Senate. Congressional Record, June 8, 1972 (daily edition): S. 9026. The study referred to is: Report on Survey of Federal Program Evaluation Practices Conducted by the Staff of Senator William V. Roth, Jr., (Delaware), May 1972: 86 pp. (Available from Senator Roth's office.) Y4EC7: B47

^{13/} Bruce, Robert G. What Goes Wrong with Evaluation and How to Prevent It. Human Needs. (Publication of Office of the Administrator, Social and Rehabilitation Service, Department of Health, Education and Welfare.) v.I, no. 1, 1972.

^{14/} Buchanan, Garth N. and Joseph S. Wholey. Federal Level Evaluation, Evaluation Fall 1972: 21-22. Weiss holds similarly that much evaluation research falls short of the expectations of funders, program staff and evaluators. See her current article: Weiss, Carol H. Between the Cup and the Lip ... Evaluation, v.I., no. 1, 1973: 49-55.

^{15/} Lewis, Frank L. and Frank G. Zarb. Federal Program Evaluation From the OMB Perspective. Public Administration Review, n. 4, July/August 1974: 314.

The literature on improving the administration of program evaluation raises important issues which may warrant additional legislative concern. These, which will be developed below, are to: clarify the purposes of evaluation; enhance communication between administrators and program evaluators; integrate program evaluation into program management; and establish better information dissemination systems.

A. The Need to Clarify Legislative Statements of Program Goals and the Purposes of Evaluation.

According to the literature, much of the failure of program evaluation can be attributed to policymakers and legislators who write vague statements of program goals, which give evaluators imprecise yardsticks or standards against which to measure program accomplishments and deficiencies or costs and benefits. Dr. Bruce succinctly characterized this lack of definition:

Program goals have to be carefully defined. There must be clear statements of what a program is supposed to accomplish -- clear in the sense that you can pick them up and go out and measure whether or not you have achieved the program goals. Too often there is a tendency to rely upon broad statements of legislative intent to describe the purpose of a program. Such statements only provide a context for defining specific and measurable program outcomes. A lack of such agreed upon definitions to serve as criteria for evaluating a program's performance may be the largest single cause of ineffective evaluation studies. ^{16/}

1. Discrepancies Between the Purposes of Evaluation and the Techniques of Evaluation. The literature reflects several variants on this theme of the need to specify program objectives and legislative requirements for program evaluation.

One theme receiving attention is: many programs are not amenable to evaluation because of the lack of correspondence between their broad objectives and the tools and techniques of evaluation research. On this point, Thomas Morehouse reports that Federal social services programs take one of two forms: maintenance programs or opportunity programs. A maintenance program "provides tangible goods or services to a clearly defined population, such as food or money. An opportunity program, in contrast, does not directly meet the material needs of a population, but is designed to "increase the group's

^{16/} Bruce, op. cit.

capabilities or opportunities to acquire goods and services (and sometimes status and power) for themselves." These opportunities include providing the individual with job training or education to give him the skills needed to obtain directly goods and services. Innovative social programs, on which most program evaluations are required, are designed to provide opportunity, not goods or services. "The consequence for evaluation," Morehouse reports, "is that program 'outputs' are often more elusive and less easily measured than the outputs of tangible maintenance benefit programs." Instead, these programs are "explorations of problems, objectives, and means." Evaluators therefore are forced to evaluate program performance by using input measures, such as manpower allocated to administer the program or funds spent. Such techniques yield very little information about the effects of the intervention strategy underlying the program. Morehouse concludes that social scientists who conduct evaluations are criticized by policymakers and administrators who "continue to call for 'objective information' and demand 'rigorous scientific' evaluation," and by fellow social scientists who call program evaluation research trivia. ^{17/}

A second widely addressed theme is that non-specific or non-quantifiable statements of legislative program objectives militate against evaluation because broad-gauged programs cannot be evaluated easily with existing social science evaluation techniques which are oriented toward experimentation. ^{18/} The experimental approach, Morehouse notes, requires that the following criteria be met: "it assumes that a program has well-defined objectives and that its effectiveness can be determined by measuring the extent to which objectives are achieved; the setting in which the program operates should be reasonably "controlled"; the program "treatment should be reasonably uniform; and the

^{17/} He adds that this is true "...even if performance (using experimentation and quantitative techniques) has often fallen far short of aspirations, and even if many evaluation research projects have been carried out in other, less highly regarded ways." (Morehouse, 871.) The basic source is Thomas A. Morehouse. Program Evaluation: Social Research Versus Public Policy. Public Administration Review [November/December 1972]: 872.

^{18/} For an explanation of social experimentation and its use in social program evaluation research, see Part V of this report.

program treatment should be applied in a large enough number of cases to provide an adequate sample of program experience." ^{19/} However, "most of the Federal social action and economic development programs to which evaluation requirements have been attached do not fit this mold." He continues:

[They often have] broad aims that cannot be specified in clear-cut form; they may be concerned more with changing a situation than with affecting a large number of individuals directly; they generally operate in open, uncontrolled community and regional settings; and a given program tends to take different forms in different situations. The programs, in short, are not themselves designed or administered as experiments. Thus it is rarely that even the basic data collection requirements for evaluation research are structured into program operations, let alone the requirements for standardized treatment, control-comparison, and representative samples. ^{20/}

2. Recommendations to Specify Program Goals in Legislation. Most evaluation in recent years constitutes a response to legislative requirements incorporated implicitly or explicitly into authorizing legislation. Four recent reports have criticized the vagueness of these legislative requirements and offered suggestions for making them more specific, to conform with the methods and procedures of current evaluation practices and to the resources authorized to conduct evaluations. ^{21/}

For example, in March of 1974, the General Accounting Office provided the Congress with an inventory and analysis of legislative language requiring program evaluation. This report, and an earlier one submitted in August of 1972, were designed to provide GAO with information to carry out its expanded evaluation responsibilities under the Legislative Reorganization Act of 1970. ^{22/} They were intended also to assist congressional committees in carrying out legislative analysis functions required under section

^{19/} Idem

^{20/} Ibid., pp. 871-872.

^{21/} GAO reports cited below; Wholey, et. al. Federal Evaluation Policy, op. cit.: 57-58; and National Advisory Council on Education Professions Development. Search For Success: Toward Policy on Educational Evaluation. A Report of the National Advisory Council on Education Professions Development, Washington, U.S. Govt. Print. Off., June 1974. pp. 35-36.

^{22/} Legislative References to Evaluation, 1967 (62 pages); preceded by a Letter from the Comptroller General to the Chairmen of each Congressional Committee, August 11, 1972. (GAO No. B-161740); and "Updated Part II of Program Evaluation: Legislative Language," in Congressional Research Support and Information Services, op. cit., pp. 338-339.

136 of the Legislative Reorganization Act of 1946, amended in 1970. ^{23/} The General Accounting Office surveyed legislative requirements for program evaluation assessing their wording along several dimensions: official responsible for evaluation, measures of effectiveness, methods required, report recipients, report dates, nature of recommendations, and funds authorized. The following conclusions were derived:

- (1) Of the 40 acts including evaluation references:
 - 15 acts are in the health and safety area;
 - 5 acts are in the education area;
 - 4 acts are in the transportation area;
 - 3 acts are in the law enforcement area;
 - 2 acts are in the housing area;
 - 2 acts are in the environmental area;
 - 2 acts are in the international area;
 - 2 acts are in the agricultural area;
 - 2 acts are in the economic opportunity area;
 - the remaining acts are each in a different area--aging, selective service, and employment.
- (2) Multiple references to evaluation (more than one section on evaluation) are included in 19 of the 40 acts.
- (3) Specific authorization for funding evaluation is included in only 6 of the 40 acts.
- (4) Specific wording concerning measures of effectiveness is included in 24 acts.
- (5) Eight acts include language specifying methods of data collection and analysis. However, each title of the Economic Opportunity Act includes very specific and detailed language on methods of evaluation including cost-benefit analysis, use of control groups, and standards for evaluation.
- (6) Specific reporting date(s) for evaluation results are included in 36 acts. Reporting requirements range from a very general time span (annually) to a very specific date (on or before January 31, for example).
- (7) There is a wide range concerning who is specified as responsible for conducting evaluations and who is specified as the recipient of the evaluation reports. Most of the laws (and sections within laws) include evaluation references specifying responsibility for conducting evaluation and dissemination of results.
- (8) Most of the laws in the health area and many of the laws in the other areas include wording for recommendations expected from the evaluations. Most recommendation language is quite specific. Recommendations are expected on a broad range of topics including changes in the legislation, changes in the program, changes in program plans, and improvement in evaluation methods and measures. ^{24/}

^{23/} According to GAO this section "directs standing committees of the House and Senate to review on a continuing basis, the application, administration, and execution of those laws, or parts of laws, within its jurisdiction in order to assist the House and Senate in (1) its analysis, appraisal, and evaluation of the application, administration, and execution of the laws enacted by the Congress, and (2) its formulation, consideration, and enactment of such modifications or changes in those laws, and of such additional legislation as may be necessary or appropriate."

^{24/} Congressional Research Support and Information Services, op. cit., pp. 338-9. The 40 acts requiring evaluation are summarized in the publication.

In support of assisting committee chairmen's efforts to improve the quality of legislative language requiring evaluation, the GAO subsequently suggested the following "model" of statutory evaluation requirements:

The (head of the agency) shall submit an evaluation report, to (committees on appropriations and committees having legislative jurisdiction over the program) no later than _____ of each year. Such report shall--

1. contain the agency's statement of specific and detailed objectives for the program or programs assisted under the provisions of this Act, and relate these objectives to those in this Act.
2. include statements of the agency's conclusions as to effectiveness of the program or programs in meeting the stated objectives, measured through the end of the preceding fiscal year.
3. make recommendations with respect to any changes or additional legislative action deemed necessary or desirable in carrying out the program or programs.
4. contain a listing identifying the principal models, analyses and studies supporting the major conclusions and recommendations, and
5. contain the agency's annual evaluation plan for the program or programs through the ensuing fiscal year for which the budget was transmitted to Congress by the President. ^{25/}

3. Criticisms of Specifying Program Goals in Legislation. It should be noted that important objections have been raised to the notion of the Congress's specifying explicit legislative program objectives and goals for program evaluation purposes. Dr. Harold Orlans, of the Brookings Institution, reports that complex political and pluralistic forces in the legislative process generate "...social programs [which] are marked by ... multiple and contending forces.... ...Congress must serve more easily by amalgam and compromise measures, which offer something to many different groups and are lent a semblance of unity by goals broad enough to embrace the spectrum of interests involved." ^{26/} He substantiates this notion, in part, with comments from Charles Schultze, based on his experiences as the Budget Bureau Director responsible for implementing the Programming-Planning-Budgeting System:

^{25/} Marvin, Keith and James L. Hedrick. GAO Helps Congress Evaluate Programs. Public Administration Review, no. 4, July/August 1974: 331.

^{26/} Dr. Orlans bases his comments on a study he prepared on "The Use of Social Research in Federal Domestic Programs," prepared for the Research and Technical Programs Subcommittee of the House Committee on Government Operations: They are in: Contracting for Knowledge: Values and Limitation of Social Science Research, Jossey-Bass, Inc., 1973. pp. 123-124.

"[Schultze] points out how diametrically opposed are the evaluator's need for a clear statement of program goals and the contrary need for vagueness in getting a program adopted in the first place:

The first rule of the successful political process is, 'Don't force a specification of goals or ends.' . . . [The] necessary agreement on particular policies can often be secured among individuals or groups who hold quite divergent ends. . . . The Elementary and Secondary Education Act of 1965 . . . was enacted precisely because it was constructed to attract the support of three groups, each with quite different ends in view. Some saw it as the beginning of a large program of Federal aid to public education. The parochial school interests saw it as the first step in providing . . . financial assistance for parochial school children. The third group saw it as an antipoverty measure. . . If there had been any attempt to secure advance agreement on a set of long-run objectives, important elements of support for the bill would have been lost, and its defeat assured. 27/

As a variant of this theme, the Research and Policy Committee of the Committee for Economic Development in a report on Improving Federal Program Performance, notes that program objectives and legislative statements of intent are so complex that adequate output or achievement measures should be developed only after the program has begun and enough time has elapsed to indicate its impacts. 28/

B. Better Communication for Effective Evaluation.

Another issue, frequently treated in the literature, is the need for improved communication between social program evaluators and agency staff who develop program objectives and requirements for evaluation, in order to enhance the correspondence between the capability of evaluators and the expectations for evaluation. Generally reports on the need for better collaboration focus on the following themes:

27/ Orlans, op. cit., pp. 124-5, citing: C.L. Schultze The Politics and Economics of Public Spending. Washington, D.C., Brookings Institution, 1968. pp. 47-49.

28/ Improving Federal Program Performance. A statement on national policy by the research and policy committee of the Committee for Economic Development, September 1971: p. 61.

-there is a need for social scientists and program administrators to determine jointly objectives of program evaluation and methodologies for research before an RFP is published; 29/

-program evaluators who are social scientists frequently prepare long, jargon-laden reports which are not written in a manner which meets policymaking needs and which are often overdue, causing Federal agencies to award funds for program evaluation to profit-making research units which prepare timely, but often inadequate and sometimes invalid program evaluations; 30/

-many social scientists shy away from doing program evaluation research because they do not consider it fruitful for generating basic information and knowledge required to advance the social sciences as disciplines, and the university peer reward structure does not value evaluation research the way it does basic research and applied research in the social sciences;

-Federal agencies are not properly organized to administer social science research programs, and administrators do not understand the complexities and limitations of social science analytical capabilities; and

29/ This issue is explored in greater detail in Part IV of this report. However it should be noted that literature on this theme is both vituperous and constructive. Reflecting the first position: Behavior Today, a newsletter distributed to the social science community, reports the following on discussion at a session of the September 1971 convention of the American Sociological Association:

NSF's Joel Snow, OE's John Evans and NIMH's Kenneth Lutterman painted a bright future for social science funding, but emphasized priority on applied and evaluation research. Evaluation research, suggested Peter Rossi from Johns Hopkins, in a breezy but blunt rebuttal, often amounts to research on nonsense. 'Sociologists ought to have the guts to say so.' Politicians, Rossi claimed, controlled the conception of the poverty program, the community health centers, which he called silly, and NSF's new Rann program which he dubbed 'a WPA for engineers and political scientists'. Sociologists, he explained to the government representatives, are hesitant to spend three years of their lives evaluating such programs. There's something wrong with social science and the federal government,' said Rossi. 'We never get a chance to define the problem. The problems, even the social scientists themselves, are defined by the reigning politicians.'

See also: Program Evaluation: Review of Major Issues and Literature, a report prepared with the Instructions of the Hon. Roman L. Hruska. By Genevieve J. Knezo, Science Policy Research Division, Congressional Research Service, Library of Congress, January 31, 1972: 247 pages. (Reproduced and distributed by Law Enforcement Assistance Administration, United States Department of Justice, May 19, 1972.)

30/ For additional information on this issue, see Part IV of this report.

-program evaluators must become more sensitive to the political considerations which impact on the initiation and use of program evaluation; they should consider evaluating the political impacts of a program or at least be sensitive to the political values impacting on a decisionmaker who may or may not use the evaluation report. 31/

31/

The theme of political/bureaucratic barriers and resistance to the use of evaluation and social experiments surfaced repeatedly during the Dartmouth/OECD Conference on Social Research and Public Policies, September 1974. In summary it was noted that the same political consensus or conditions which permit the conduct of experiments or evaluations produce independent political consensus to adopt a program even before the results of the research are available. Similarly, an evaluation or experiment which generates positive results for the intervention strategy will not lead to large-scale adoption of the program if political conditions do not permit it. Good evaluations which yield negative results for program effectiveness may not be used because legislators and agency program managers don't want to see their pet project cut. Good or bad evaluations which yield positive results for program effectiveness may spell the demise of a program since the program may be perceived as having achieved its purposes.

For additional information on these topics see: Campbell, Donald T. *Assessing the Impact of Planned Social Change*, Background Paper for the Dartmouth OECD Seminar on Social Research and Public Policies, September 13-15, 1974, passim; Weiss, Carol H. *Where Politics and Evaluation Research Meet*. *Evaluation*, v. 1, n. 3, 1973: 37-45; and Boeckmann, Margaret. *Policy Impacts of the New Jersey Income Maintenance Experiment*. A paper presented at the American Sociological Association Meetings, Montreal, August 25-29, 1974, mimeo: 28 pp.

An in-depth study on how decisionmakers use and do not use scientific and technical information, including social science information, is being conducted by Nathan Caplan at the Institute for Social Research University of Michigan. For a short description, see: *Science Is Seldom Put to Good Use by U.S. Officials*, LSR Newsletter, v. 2., n. 1, Spring 1974: 2, 8.

Recommendations for social scientists to devote more attention to the broker concept in communications -- to obtain a better awareness of policymakers' needs and modes of behavior, decisionmaking requirements, and constraints may be found in: Lane, Robert. *Social Science Research and Public Policy*. *Policy Studies Journal*, v. 1, n. 2, Winter 1972: 111; Jones, Ernest. *Law, Political Science, and Policy Studies*. *Policy Studies Journal*, Autumn 1973: 56-71; Elkin Stephen L. *Political Science and the Analysis of Public Policy*. *Public Policy*, v. 22, n. 3, Summer 1974: 399-422; Shonfield, Andrew. *The Social Sciences in the Great Debate on Science Policy*. *Minerva*, July 1972: 426-438; Horowitz, Irving Louis. *Social Science Mandarins: Policymaking as a Political Formula*. *Policy Studies*, v. 1, Fall 1970: 339-360; and Levin, Martin A. and Horst D. Dornbusch. *Pure and Policy Social Science: Evaluation of Policies in Criminal Justice and Education*. *Public Policy*, v. 21, n. 3, Summer 1973: 383.

C. The Integration of Program Evaluation Into Program Management.

Recent congressional reviews of program evaluation in specific agencies and social science assessments of ways to improve the use of program evaluation research stress that evaluation would be more useful if these activities were better integrated into policy planning and program management. It is asserted that social scientists, program evaluation researchers, and program evaluation managers must take steps to enhance this cross-fertilization of the conduct and utility of such research.

1. Incorporation of Evaluation Into Agency Program Planning and Budgeting. The absence of a link between evaluation research and decisionmaking, especially budgeting, is probably the most important impediment to effective use of evaluation research. Buchanan, Horst and Scanlon recently completed an examination of the evaluation and decisionmaking systems of the Department of Labor and the National Institute of Mental Health. On the need to incorporate evaluation into decisionmaking in these agencies, and by implication into others, they observe:

[An organizational adjustment is needed to establish] a formal relationship between evaluation planning and an agency's decision processes. In the absence of such a relationship, the same types of evaluation questions are raised year after year in different decision contexts -- issue papers, strategy papers, and legislative proposals -- without any apparent contribution or resolution by the evaluation system. If the evaluation information is ever to penetrate agency operations, the evaluation models will have to become an integral part of the decision process. This will require training of key staff -- evaluation planners, program managers, policy-makers -- in the design and use of the evaluation models. 32/

Buchanan and Wholey observe that Federal agencies have a responsibility for incorporating evaluation into program management systems and that "the establishment of an evaluation system usually requires the integration of several systems relating to planning, management, and data collection." 33/ Such integration necessitates

32/ Buchanan, Garth Pamela Horst, and John Scanlon. *Improving Federal Evaluation Planning*. *Evaluation*, v. 1, no. 2, 1973: 90.

33/ Buchanan, Garth N. and Joseph S. Wholey. *Federal Level Evaluation*. *Evaluation*, Fall 1972: 22 and passim.

"coordination among several organizational units and sometimes ... changes in the administrative structures themselves," actions which are not easily taken by entrenched Washington bureaucrats. Walter Williams, a former chief of the Research and Plans Division, Office of Economic Opportunity, evaluated the organizational problems surrounding the failure of major social program evaluations of HEW, HUD, the Department of Labor and the Office of Economic Opportunity in the late 1960s. ^{34/} He made the following administrative recommendations to fashion closer functional authority between research and implementation: (paraphrase)

- (1) the central research analyst should be given more power to generate research requirements and policy analysis required to develop recommendations regarding program funding and implementation;
- (2) the analyst should be given more money to conduct long-range social experimentation research and evaluation;
- (3) more feedback loops should be built between the field and the central office to overcome some problems of implementation;
- (4) the central analyst should have capability to fund social science research which leads to recommendations on program improvement rather than merely program achievement; and
- (5) more attention should be paid to stressing outcome objectives in implementation rather than input measures. ^{35/}

^{34/} Williams' major points were summarized in an article reviewing major recent critiques of program evaluation and social experimentation methodology: "(1) the major failure of social action programs in the 1960s was in not being able to move from the design of the program to the implementation of the program, particularly in the field; (2) the major deficiency in data was that it was macronegative (showing how many blacks were below the white age group in certain reading achievement tests instead of micropositive (how do you improve that reading level)); (3) the central analysis focused too much on what was happening in Washington and not enough on what was happening at the local level; and (4) the social science research community is not structured to respond to the analytical needs of social agencies - its reward structure mitigates against research directly in support of policy analysis." (Robert E.C. Wegner. Evaluating the Evaluators. Public Administration Review [January/February 1973]: 85-6).

^{35/} Williams, Walter. Social Policy Research and Analysis: The Experience in the Federal Social Agencies. New York, American Elsevier Publishing Company, Inc. [1971] pp. 180-1.

Weiss's report on administrative obstacles to effective program evaluation in the National Institute of Mental Health concludes:

"...While not all of the impediments can be planned away in advance, certain basic [organizational] conditions should be present before evaluation grants are

2. The Development of Evaluation Work Plans. A second class of problems discussed in relation to incorporating program evaluation into management is the need for an agency to develop comprehensive work plans for evaluation. This includes setting priorities among programs to be evaluated and designing a succession of evaluation studies which provide answers required to determine the adequacy of the program over time. Wholey, et. al. summarized this problem in the 1970 study:

With few exceptions, Federal agencies have not had adequate work plans for evaluating their major social programs.

The lack of evaluation work plans often resulted in haphazard letting of unrelated contracts to evaluate the same program. Contractors designed their own studies, formulated their own questions, and naturally produced data and findings that were not comparable with each other. ^{36/}

3. LEAA: An Illustration of These Deficiencies. A congressional investigation of the administration of the block grant program of the Law Enforcement Assistance Administration attributes major failures in program planning and evaluation to the absence of a well-formulated plan for evaluation. Many of the findings of this study probably also characterize the social program evaluation mechanisms of other Federal agencies. Because of their generality and potential contribution to additional legislative inquiry on evaluation mechanisms, these findings will be summarized:

-Although the LEAA has authority under its enabling legislation to "conduct evaluation studies of the programs and activities assisted under the title," the agency has done little toward making its own evaluation of the effectiveness of programs or projects funded with block grant funds. ^{37/}

-One of the major deficiencies LEAA encounters in developing adequate evaluation plans is that because of poor information systems, the agency lacks knowledge

made. Among the most important are: (1) clarity of purpose for the study, (2) a well-defined and relatively stable program, (3) administrative support, and (4) an able research staff who give a substantial proportion of time to the study." (Weiss, Carol H. Between the Cup and the Lip... Evaluation, v. I, no. 2, 1973: 55.)

^{36/} Wholey, et. al., Federal Evaluation Policy, op. cit. p. 35.

^{37/} U.S. Congress. House Committee on Government Operations. The Block Grant Program of the Law Enforcement Assistance Administration. Part 1, Hearings before a Subcommittee of . . . July 1971. Washington, U.S. Govt. Print. Off., 1971. p. 143.

about the disbursement of grant funds and the purposes to which these funds will be spent. This deficiency should be rectified with an improved grant management information system. 38/

-Equally, if not more important in terms of evaluating management performance and capabilities, neither LEAA nor State agencies which administer the programs have formulated standards for evaluating program progress, success, or failure. Thus, the programs are unevaluated, unaudited, and incapable of being measured as to performance and progress due to the lack of goals or standards. 39/

-Since LEAA does not maintain an inventory of grant programs and lacks a centralized evaluation plan and standard, its efforts sometimes duplicate those of other agencies. 40/

-Many LEAA funds are administered directly by State Planning Agencies, (SPA). The lack of program goals and evaluation standards and guidelines prevents SPAs from adequately administering and evaluating programs. LEAA has the responsibility for providing SPAs with manuals and guidelines for evaluation but has not adequately responded to this need. 41/

4. The Contribution of Evaluation Work Plans to Evaluations Performed at the State Level. Federal assistance to State and local jurisdictions for developing evaluation plans, objectives, methodologies, and information-dissemination systems seems all

38/ Specifically: ". . . LEAA [lacks] knowledge of the purposes, goals, and specifications of the thousands of State and local projects financed by its block grants. This informational void, still unfilled as LEAA completes its fourth year of operation, makes LEAA powerless to modify programs to avoid deviations and to follow congressional intent. A grant management information system and a reference service, both of which should alleviate these and other monitoring problems of LEAA, are currently being devised by outside consultants under contract to the agency." (U.S. Congress. House. Committee on Government Operations. Block Grant Program of The Law Enforcement Assistance Administration. Twelfth Report by the Comm. on Government Operations. House Report. No.92-1072. May 18, 1972. Washington, U.S. Govt. Print. Off., 1972. p. 8.)

39/ Idem.

40/ Ibid., p. 61.

41/ Specifically: Statement of Mr. Ahart. "In the SPA Guide for 1969 (a manual issued by LEAA to the State Planning agencies for guidance on application awards, and administration of planning and block grants). LEAA stated that it would issue guidelines suggesting appropriate procedures, techniques, and measures for evaluating the contribution to crime control of the block grant projects and expenditures. The guide provided that the State planning agencies, pending issuance of the guidelines, outline in their 1969 State plans a tentative program for project evaluation and the measurement of overall plan performance. . . . We were advised by the LEAA that it had not issued guidelines to the State planning agencies on evaluation methods because of a shortage of manpower. Also we noted that on occasions, information or guidance has been requested from LEAA on monitoring and evaluation methods, and LEAA has been unable to provide the assistance." (The Block Grant Programs of the Law Enforcement Assistance Administration, (Part 1), hearings, op. cit., pp. 143-4

the more imperative since State and local governmental units have been given greater responsibility for program administration and accountability under revenue sharing. A few agencies have encouraged developments in this area, notably Housing and Urban Development and the Office of Management and Budget. 42/ In a partial effort to assist non-Federal authorities in performing required audit functions, the General Accounting Office prepared a report in June 1972, titled Standards for Audit of Governmental Organizations, Programs, Activities, and Functions, 54 pages. This report does not however, provide standards for program evaluation. The GAO is now preparing a program evaluation guidebook, expected to be released in the Spring 1975.

5. Proposals for Comprehensive Evaluation Plans. A number of agencies conducting program evaluation research have begun to develop coherent evaluation work plans to guide both internal agency evaluators and extramural evaluators. 43/ As will be noted below, the Office of Management and Budget is now encouraging activities in this direction. It may be of interest to the Congress to inquire into whether other Federal agencies are taking steps to provide for systematic development and use of evaluation work plans. Wholey, et. al. made a number of recommendations treating this issue. Some of their observations are reproduced below:

42/ See: Ninth Conference on Management Analysis in State and Local Government, November 2 and 3, 1972, sponsored by Institute of Public Service, University of Connecticut. Among the papers were: Program Evaluation: Pitfalls and Potentialities, Dr. Jack Carlson, OMB; Case History Application of Program Evaluation, Keith Marvin, General Accounting Office. The most useful study in this respect is the HUD funded report: Hatry, Harry P. et. al. Practical Program Evaluation for State and Local Government Officials. Washington, D.C., The Urban Institute, September 1972. 134 pp.

43/ For an introduction to this topic, see R.A. Levine and A.P. Williams, Jr., Making Evaluation Effective: A Guide. A report prepared for the Department of Health, Education, and Welfare. Santa Monica, The RAND Corporation, May 1971, 40 pp. (R-788-HEW/CMU.) Fred Spaner and Charles Winder, National Institutes of Mental Health. The Perspective of National Program Development and Administration: Program Needs and Evaluation Activities. Unpublished, February 5, 1971. Presented at American Psychological Association Convention September 1971. Quotation from Federal Evaluation Policy, op. cit., pp. 114-115.

EVALUATION PLAN

At the beginning of each fiscal year, each agency should develop or update a two to three-year evaluation work plan stating which studies are in progress and which are projected. Approval of this plan should be required at the beginning of each fiscal year as a condition for authority to spend evaluation funds. This plan should be prepared by the agency-level evaluation staff in cooperation with policy makers, budget staff, program managers and operating-level evaluation staffs. Evaluation planning is a continual process: when the agency appropriation is known, the evaluation plan should be reassessed in light of firm budget figures.

Since useful evaluation studies are likely to be costly, agencies should place emphasis on feasible studies of major programs where the value of the findings would outweigh the costs. ^{44/}

6. Proposals for a "Preassessment of the Evaluability of Programs." As a variant of this theme, researchers at the Urban Institute suggest that the conduct and use of program evaluation will be enhanced if federal agency administrators and program evaluation researchers conduct a "preassessment of the evaluability of a public program." The objective of the assessment would be to select appropriate research strategies for evaluation. The following questions, addressing the assumptions of the program, its data base, and the probable use of the evaluation in decisionmaking processes, suggest a sequence of evaluation steps and criteria to be used in determining the "evaluability of a public program" and priorities among different types of evaluation strategies:

- Are the problems, intended program interventions, anticipated outcomes, and the expected impact sufficiently well defined as to be measurable?
- In the assumptions linking expenditure to implementation or intervention, intervention to the outcome anticipated, and immediate outcome to the expected impact on the problem, is the logic laid out clearly enough to be tested?
- Is there anyone clearly in charge of the program? Who? What are the constraints on his ability to act? What range of actions might he reasonably take or consider as a result of various possible evaluation findings about the measures and assumptions discussed above? ^{45/}

^{44/} Federal Evaluation Policy, op. cit., pp. 114-5.

^{45/} Horst, Pamela, Joe N. Nay, John W. Scanlon, and Joseph S. Wholey. Program Management and the Federal Evaluator. Public Administration Review, v. 4, July/August 1974: 307. (Part of PAR/special "Symposium on Program Evaluation.") (Footnote continued on p. 23.)

D. The Need for Better Dissemination of Information on Program Evaluations Supported Under Federal Agency Auspices

Several issues are subsumed under the general heading of the need for improved program evaluation information dissemination systems. Among them are: the need for researchers to share experiences in developing methodologies to hasten refinement of program evaluation techniques; to eliminate duplicate funding of methodologies and evaluations of programs; and to insure distribution of program evaluation reports to individuals within an agency and to individuals outside of the agency whose program responsibilities intersect with the program being evaluated.

Howard R. Davis, chief of the Mental Health Services Research and Development Branch of the National Institute of Mental Health, has summed up these concerns:

Increasing numbers of investigators are working vigorously on methods of measurement and analysis in evaluation; but one researcher hardly knows what the others are doing. Duplication unavoidably is occurring. Incompatible counterparts of evaluation systems are needlessly re-invented. . . . Better methods of exchanging information [should] be developed as soon as possible. Not only would each person producing evaluation knowledge know what others are doing, not only would those needing to implement evaluation programs have a valuable perspective of what is going on elsewhere, but the concepts of evaluation. . . would receive continued clarification. . . . ^{46/}

Perhaps as a partial response to this need, the National Institute of Mental Health awarded a grant to the Program Evaluation Project, Minneapolis Medical Research Foundation, to establish, on an experimental basis, Evaluation, the journal in which this statement is found.

The National Advisory Council on Educational Professions Development made a related recommendation for the administration of each specific program evaluation: "Evaluation research plans can fail in implementation just as programs can fail. Time must be taken to 'walk through' evaluation research designs by applying them to simulated program data, operational settings and decision questions, before applying the evaluation research design to the actual program. Such a process could create much more understanding of various models of evaluation research and alert decision-makers to potential problems and results in advance." (Search for success. . . ., op. cit., 24.)

^{46/} David, Howard R. Insights: A Solution for Crisis? Evaluation, v. I, no. 1, Fall 1972: 3-4.

1. Investigation by the General Accounting Office. A number of recent reports indicate that far more systematic methods of exchanging information on program evaluation may be required. For example, the General Accounting Office published a report entitled "Program Evaluation: Legislative Language and a User's Guide to Selected Sources," in June 1973. This report was prepared to assist GAO researchers obtain published information and program evaluations to which they could refer in performing their auditing and evaluation functions for the Congress. The author describes and assesses 30 on-going research inventory and published information sources established by Federal agencies. These, among others, include the National Technical Information Service (NTIS); the Science Information Exchange (SIE); ERIC, a research report inventory for federally funded education research; and MEDLARS and NCHSRD (inventories of research in progress or published related to medical and mental health services, respectively).

The report indicates the following inadequacies:

Despite [the] rapid growth in evaluation studies, information as to who is conducting which studies with what findings is not readily available. No single-source document or retrieval system identifies planned, ongoing, and completed evaluations. No single source describes the findings and conclusions of completed evaluations. In fact, no single functional area (such as health, education, or training) has such complete coverage of evaluation studies. Since program evaluation provides perhaps the most comprehensive, objective, and quantitative sources of information on governmental program results, it is important to the government and to the public that these results be identified and made available. Most non-defense evaluations are public information and need to be listed and described.

A number of computer retrieval systems and reference documents listing a large number of studies, including evaluation studies, are available. However, knowledge is generally lacking as to which information sources exist, which sources contain evaluation studies, how comprehensively evaluation studies are described, and how to obtain information from such sources. This chaotic situation makes it a long-term task to develop, maintain, and distribute a comprehensive list and description of planned, ongoing, and completed evaluation studies in all functional areas. ^{47/}

2. NIMH/UCLA Project on Data Bank of Program Evaluation (DOPE). It should be noted that the Health Services and Mental Health Development Branch of the National

^{47/} Program Evaluation: Legislative Language and a User's Guide to Selected Sources, op. cit., p. 1.

Institute of Mental Health is funding a joint project with UCLA to establish an easily accessible computerized information system on program evaluation, called the Databank of Program Evaluations (DOPE). The system, expected to be in operation in early 1975, will enable all researchers who have access to the ARPA computer network (of the Advanced Research Projects Agency) to obtain an abstract of the following information for each program evaluation report and secondary source document included in the system:

- citation,
- status of the program success,
- condition treated
- age, sex, race and income characteristics of the sample population,
- sample size,
- site,
- treatment method,
- characteristics of the study design,
- what is measured,
- what measures are used,
- how data is collected, and
- conclusions.

Evaluation reports will encompass the following topics: alcohol abuse, divorce, juvenile delinquency, neuroses, sexual deviancy, suicide, welfare, health problems with mental health outcomes, criminality, drug abuse, mental retardation, psychoses, sociocultural problems, unemployment, and mental problems of the aged. Three types of search are being used to identify reports of evaluations: (1) journal articles, (2) published and unpublished reports recommended by experts in each subject field, and (3) materials from other computerized data banks, such as Medlars, Psychological Abstracts, and Sociological Abstracts. ^{48/}

^{48/} Wilner, Daniel M., Robert W. Hetherington, Ellen B. Gold, Daniel H. Ershoff, and Cedric F. Garagliano. Databank of Program Evaluations. Evaluation, v. I, no. 3, 1973: 3-6. (The project is supported under grant IR 12-MH-2276.)

PART III. IMPROVED COORDINATION OF
PROGRAM EVALUATION IN THE EXECUTIVE
AND LEGISLATIVE BRANCHES

A. Issues Surrounding the Need to Incorporate Program Evaluation Findings Into
Executive Branch Budget Processes

Two topics are included under this heading of improved coordination of program evaluation in the executive branch. The first relates to ways to further incorporate program evaluation findings into executive branch budget processes. The second deals with improved coordination of the administration and use of program evaluation in the executive branch.

The only government-wide requirements for agencies to incorporate program evaluation data into annual agency budget proposals are found in Circular A-11, issued by the Office of Management and Budget. The term "program evaluation" is used very loosely in this circular, since costs and benefits or program accomplishments or outputs are to be stated directly in dollar terms or in number of persons affected, rather than in more hard to measure terms of real, but mostly intangible, program accomplishments. For instance, data to be submitted for Federal Income Security Programs consist of three types: benefit outlays, number of beneficiaries and average monthly benefits, number of individuals benefited by the outlays and number of primary beneficiaries. These data are required of separate target groups categorized as aged or retired; disabled; mothers; surviving spouse; other survivors; unemployed and others. (Sec. 49.) These data requirements differ considerably from requirements for most agency program evaluations, which are designed to determine whether a program should be modified, terminated, or continued. Evaluation requirements under Circular A-11 also do not yield information to determine which among alternative social programs best serves a particular target group or meets an accepted national need. These requirements have also been widely criticized as inappropriate

to the budgeting process in both executive and legislative branches since almost no attempt is made by the Office of Management and Budget to integrate the results of program evaluations and program evaluation information to the annual budget process.^{49/} Criticism has been heard also that what information the executive branch does collect is withheld from the Congress all too frequently on the grounds of executive privilege. Commenting on this point in 1970, (when Office of Management and Budget was called the Bureau of the Budget), the Joint Economic Committee said:

Program Memoranda, containing quantitative evaluation of the benefits and the costs of individual programs as well as other evaluative information, have been retained ... on grounds of executive privilege. Special Analytic Studies, presenting the results of detailed study of the efficiency and equity consequences of alternative means of obtaining objectives, are from time to time released, but only after the executive has made a policy decision and then only in a form designed not to upset the decision. Program and Financial Plans, with program budget projections, are held within the executive on grounds of lack of data and the unreliability of estimates. While the list of issues on which the Bureau of the Budget is requesting analysis in the planning and budgeting cycle has been released to the Congress through the subcommittee, the results of the analyses of these issues have been withheld ...

... The lack of congressional access to much of this information and analysis seriously hinders legislative efforts to scrutinize and control the budget. The widespread use of executive privilege to retain the results of program evaluation and analysis is inimical to the development of an open and responsive political system in which public policy is made by elected political leaders rather than by an anonymous and powerful bureaucracy.^{50/}

And although OMB asks agencies to describe generally how programs meet national objectives, termination of requirements for agencies to adhere to the Planning, Programming, Budgeting system, in effect, removed requirements for agencies to analyze

^{49/} Allen Schick, "An Analysis of Proposals to Improve Congressional Control of Spending." In U.S. Congress. Joint Economic Committee. Subcommittee on Budgeting, Management and Expenditures. Improving Congressional Control Over the Budget: A Compendium of Materials. March 27, 1973. Washington, U.S. Govt. Print. Off., 1973. pp. 223-4. (Committee Print.)

^{50/} U.S. Congress. Joint Economic Committee. Subcommittee on Economy in Government. Economic Analysis and the Efficiency of Government. Report. February 9, 1970. Washington, D.C., U.S. Govt. Print. Off., 1970. p. 9. (Joint Committee Print.)

program accomplishments and deficiencies in relation to program planning.^{51/}

^{51/} General information on removal of requirements for PPB, stemming from administrative problems and difficulties in obtaining required information, are reviewed in: Schick, Allen. A Death in the Bureaucracy: The Demise of Federal PPB Public Administration Review, vol. 33, March/April, 1973: 146-156. Specific details on requirements for submitting evaluation data in budget proposals and on the evolution of PPB requirements were supplied by Mr. Joel Anderson, Office of Management and Budget:

According to Mr. Anderson the most important OMB directive on program evaluation is Bulletin No. 68-9, issued April 12, 1968, supplemented specifically by Bulletin No. 68-9, supplement No. 1, July 17, 1969, and supplemented generally by Circular No. A-11, usually revised and re-issued each year. Bulletin No. 68-9 and supplement No. 1 treat the Planning-Programming-Budgeting System. Circulars A-11 treat preparation and submission of annual budget estimates.

These OMB directives treat program evaluation only very generally within the context of "programming" as part budgeting/appropriations submissions to the Congress. Effectiveness is treated in this sense from an economic approach. The "programs" which are to be evaluated for the purposes of preparing data required by these directives are general budget categories, not specific operational agency programs.

President Richard M. Nixon issued a Memorandum to the heads of Federal departments and agencies, May 25, 1970, in which he outlined the importance of evaluating effectiveness and using program evaluation to underpin agency planning and budgeting activities. He outlined general criteria which the required program evaluations were to meet. OMB circular A-11, 1971 (With transmittal Memorandum No. 38), reflects the need for agencies to meet these general criteria in preparing budget estimates. However, the criteria outlined in the Circular are discretionary and diverge somewhat from the specific requirements under the rescinded PPB guidelines. ("Analysis of Changes in Office of Management and Budget Circular No. A-11, Revised June 1971, Comparison of Old Sec. No. 24.1-24.6 to 24.1-24.5, p. 3.)

Program evaluation requirements in the current circular, A-11, Revised, Transmittal Memorandum No. 39, June 12, 1972, are similar to the generalized requirements of the 1971 circular. These requirements are outlined in paragraphs 24.2-24.3 (Analysis and Justification of Programs' and 36.1-2 Narrative Statements on Program Performance.) The material requested to fulfill requirements for "Analysis and Justification of Programs" supports the first part of the budget request; material required to meet "Narrative Statements on Program and Performance" in the circular is used for detailed budget appendix. Mention is made of a number of techniques for conducting evaluations, and for relating program objectives and inputs to program benefits and social benefits. However, the Circular seems to leave it up to the discretion of the agency director to use techniques which are most appropriate in evaluating the effectiveness of its programs.

Specific "evaluation" requirements for programs falling under separate functional areas, such as manpower, health, and education, are also detailed in Circular A-11.

B. Proposals to Give the Office of Management and Budget Program Evaluation Responsibilities

A number of proposals have been made to give the Office of Management and Budget explicit coordinating and governing responsibilities for program evaluations conducted by agencies of the Government. Joseph S. Wholey, et. al., examined this issue in a 1970 report which analyzed the evaluation activities of four Federal agencies. They reported: "The executive branch has no system for identifying major questions to be answered through evaluation, for ensuring that high priority studies were carried out, or for ensuring that significant evaluation findings reached appropriate decisionmakers." They recommended that: the Bureau of the Budget (now called OMB) (a) require agencies to develop two-to-three year evaluation plans and to use evaluation information in budgetmaking and legislative review; (b) evaluate the quality of evaluation techniques used by different agencies; (c) conduct cross-agency evaluations "to compare the effectiveness of related Federal programs in achieving common objectives;" and (d) enlarge budget allocations for evaluation staff and resources.^{52/}

C. The Proposal to Establish a Federal Program Evaluation Office in Either the Office of Management and Budget or in the Executive Office of the President

Laurence E. Lynn, Jr., who served as Assistant Secretary for Planning and Evaluation in the Department of Health, Education and Welfare, between 1971 and early 1973, made the most recent proposal in this respect. In order to overcome

^{52/} Federal Evaluation Policy, op. cit., 24, 58-61. Similar recommendations have been made by the Committee for Economic Development in its report Improving Federal Program Evaluation, op. cit., 59 and by the President's Advisory Council on Executive Organization. Memorandum for the President of the United States on Establishment of a Department of Natural Resources [and] Organization for Social and Economic Programs (including background memoranda and attachments). February 5, 1971, Appendix 5.

the many inadequacies of current program evaluation practices, he proposed the creation of a Federal Program Evaluation Office in either the office of Management and Budget or in the Executive Office of the President to "[supervise] the evaluation activities of federal program agencies, [to conduct] a program of research on evaluation methods, and [to produce] an annual evaluation report."^{53/} Specifically the proposed Federal Program Evaluation Office (FPEO) would:

... determine how to apportion the funds among the FPEO and the agencies in accordance with a comprehensive plan

... be directly responsible for, and retain full authority over, all program impact evaluations. Adequate qualified staff would be maintained to do the necessary planning, write requests for proposals, review proposals, and most important, closely monitor the work of the evaluation contractor or grantee.

... be responsible for overseeing the development of plans for program strategy evaluations and project evaluations

... prepare the annual evaluation report summarizing evaluation findings and work in progress

... operate a system for rating the performance of evaluation contractors and grantees and for making the ratings available to evaluation officials.

... [develop a priority framework] to guide the allocation of evaluation resources ... in recognition of resource limitations, the potential value of possible valuations, and the availability of evaluation capacity.^{54/}

D. Current OMB Initiatives in Program Evaluation

As noted above, the OMB has stepped up its evaluation activities in order to improve the management and use of evaluation research. In a recent article two OMB officials, who were then responsible for directing the agency's program evaluation activities, outlined the specific objectives of the present OMB effort. These objectives, directed at each Federal agency, are:

^{53/} Lynn, Laurence E. Jr. A Federal Evaluation Office? *Evaluation*, v. I, no. 2, 1973: 29, 62.

^{54/} *Ibid.*, pp. 62,96.

(a) development of built-in evaluation capability from the inception of each new Federal program; (b) adequate staff to assure quality of output; (c) sufficient planning to focus attention on the high pay-off areas; (d) follow-up procedures to assure that results are utilized (e) organizational alignments to assure full integration with the policy decision process and related and supportive activities; and (f) a proper balancing of long-term and short-term efforts (in-house and external) to meet both quality and timeliness needs.^{55/}

Briefly, OMB's evaluation activities consist of the following: (1) establishment of a Division of Evaluation and Program Implementation under the Associate Director for Management and Operations; the Office is directed by Mr. Clifford W. Graves; (2) establishment "on a very informal basis of a panel of senior Federal executives to serve as a forum for the mutual exchange of ideas and developments within the Federal evaluation community;" (3) initiation of a "... comprehensive survey of Federal evaluation programs that will inventory and catalogue ... such activities;" (4) development of a "cross-agency budget review to assess the adequacy of funding levels for evaluation programs in a number of Federal agencies;" (5) encouragement of all agencies to prepare "formal plans on a periodic basis to direct and guide individual [evaluation] studies to high priority topics;" (6) review of legislative initiatives "to assure that adequate provision is made for evaluation" and for establishment of an appropriate evaluation capability for the program; (7) incorporation of evaluation priorities into the management by objectives effort; and (8) monitoring, on a selective basis, of evaluation studies of particular interest to OMB and the White House.^{56/}

As a first step in carrying out these responsibilities, the Evaluation and Program Implementation Division of OMB conducted a survey of federal agency

^{55/} Federal Program Evaluation From the OMB Perspective, op. cit., p. 317; each of these objectives is fully described on pp. 314-317.

^{56/} *Ibid.*, pp. 312-314.

evaluation activities which provides information on the following points:

- persons(s) in the agency responsible for performing different types of evaluations;

- types of activities undertaken;

- magnitude of resources available for evaluation;

- characterization of relationships between the production of evaluation information and decision processes such as management by objectives, budget preparation, and statistical planning; and

- persons who can provide additional information on evaluations conducted.

A draft version of the report was released in June. It is expected that the materials will be updated and disseminated more widely after additional agencies provide responses to the OMB questionnaire.^{57/}

E. Issues Relating to the Improvement of Legislative Branch Program Evaluation Capabilities

The absence of a congressional program evaluation capability has repeatedly been cited as a major impediment to incorporating program evaluation reports into the legislative process. Two frequently mentioned themes are: (1) that the executive branch both conducts program evaluations and assesses the results of program evaluations without presenting reports to the Congress or to an objective critic,^{58/}

^{57/} Survey of Federal Evaluation Activities. By Evaluation and Program Implementation Staff, Office of Management and Budget, June 1974, draft, loose leaf, various pagings.

^{58/} On this point Schick reports: "... Evaluation comes too late, after the program has garnered strong client support and is built into the budget base. The classic case is summer Headstart which was serving hundreds of thousands of preschoolers when adverse findings were released. Popular programs are hard to terminate, no matter what judgments flow from the evaluations." ("The Pilot Testing of New Programs: An Analysis of Title IV, S. 3984," In: Improving Congressional Control of the Budget, op. cit., p. 278)

and (2) the Congress does not have necessary staff expertise to analyze such reports even if they were provided. Murray L. Weidenbaum, et. al. summarized these problems:

Evaluative research is now a virtual monopoly of the executive branch, to the extent that it is used at all. If Congress is ever to oversee executive administration effectively, then it needs professional assistance beyond that provided by existing committee staffs. As imperfect as evaluation techniques are, at the least they would be a useful supplement to existing rules-of-thumb and other subjective approaches.^{59/}

In this same respect the Joint Economic Committee reported:

The relationship of the executive and legislative branches pertaining to the use of executive analysis and evaluation documents is a serious problem. More serious, however, is the failure of the Congress to provide itself with an analytical capability. Currently, the Congress has neither an adequate capability to interpret or evaluate studies done by the executive and those outside of Government nor the necessary capacity to undertake policy analysis of its own.^{60/}

Evaluation of public welfare programs would be enhanced, the Joint Economic Committee reported in November 1972, by creating an evaluation agency independent of the administering agency and reporting directly to the Congress:

... Enabling legislation usually assigns the task of evaluation to the program administrators. Separation of powers is a well-accepted and venerable principle, and its application is as appropriate here as elsewhere. Administrators are understandably anxious to depict their programs as successful, and evaluations conducted by them (no matter how conscientious they may be) cannot escape being suspected of bias. An independent agency accountable to Congress, should be responsible for evaluation.^{61/}

^{59/} Weidenbaum, Murray L., Dan Larkins and Philip N. Marcus. Matching Needs and Resources: Reforming the Federal Budget. Washington, D.C., American Enterprise Institute for Public Policy Research [1973] p. 11.

^{60/} U.S. Congress. Joint Economic Committee. Studies in Public Welfare, Paper No. 3, The Effectiveness of Manpower Training Programs: A Review of Research on the Impact of the Poor. A staff study prepared for the use of the Subcomm. on Fiscal Policy. November 20, 1972. Washington, U.S. Govt. Print. Off., 1972. p. 13. (Joint Committee Print.) The other studies in the series are: Paper no. 1, Public Income Transfer Programs: The Incidence of Multiple Benefits and the Issues Raised by Their Receipt, April 10, 1972; Paper No. 2, Handbook of Public Income Transfer Program: How They Tax the Poor, December 22, 1972; Paper No. 5, Part 1. Issues in Welfare Administration: Welfare---An Administrative Nightmare, December 31, 1972.

^{61/} The Effectiveness of Manpower Training Programs: A Review of Research on the Impact of the Poor, op. cit., p. 15.

Within the last year or two proposals were made to provide the Congress with a program evaluation capability independent of the executive branch. Most of these responded to alternative ways of meeting some of the recommendations of the Joint Study Committee on Budget Control.^{62/} One of the committee's major observations was that legislative determination of priorities among programs and of spending levels for these programs would be strengthened if the Congress coordinated more closely the appropriations process and the authorization process (which now fragments legislative review of the evaluation of the effectiveness of particular programs, evaluation of the merits of one program against another, and program funding).^{63/}

During the 93rd Congress, both Houses directed attention to improving the process of determining and evaluating budget priorities to serve national needs and to strengthening the review of the budget process and evaluation of Federal programs. The Congressional Budget and Impoundment Control Act, P.L. 93-344, July 23, 1974, includes two sections which expand legislative program evaluation capabilities. Under section 701, committees are given the authority to conduct evaluations themselves, using pilot tests and cost/benefit analyses, if warranted, or to require agencies to evaluate programs and report the results to committees. Section 702 expands the review and evaluation functions of the General Accounting Office. Under this section the GAO is charged with responsibility for assisting committees in evaluating government programs, including the development of statements of legislative objectives,

^{62/} The committee was created in October 1972, pursuant to P.L. 92-599, to recommend procedures for improved congressional control over the outlay and receipt totals in the budget and to help establish spending priorities in the budget. The committee presented its interim report to the Congress on February 6, 1973, "Improved Congressional Control Over Budgetary Outlay and Receipt Totals."

^{63/} See section II. "Tentative Recommendations. Introductory paragraph." (In Interim Report by the Joint Study Committee on Budget Control. Remarks of Hon. Mr. Ullman on the floor of the House. Congressional Record, February 7, 1973: H. 843.)

methods for review and evaluation of programs, and the analysis of program results. The Act also authorizes the establishment of an Office of Program Review and Evaluation in the General Accounting Office.^{64/}

Some of the other bills considered during the 93rd Congress included provisions which would have made pilot testing mandatory or would have required OMB to inventory or to use program evaluation information in preparing budget submittals. (S. 40, Mr. Brock; S. 1030, Mr. Humphrey; S. 758, Mr. Beall; and S. 834, Mr. Muskie. S. 758, in particular, would have created a Federal program evaluation digest.)

Related to this was a House proposal, H. Con. Res. 21, introduced by Mr. Conte and referred to the Committee on Rules, which would establish a "Joint Select Committee on Government Program Analysis and Evaluation." The Committee would be authorized to evaluate alternative procedures and to evaluate "the efficiency and effectiveness of Federal programs and activities by objective, scientific, and empirical analysis...." Among the alternative proposals the Committee would examine are:

- "the establishment of an ad hoc, independent, bipartisan commission to study and appraise Federal programs and activities...of the executive branch...;"

- "an expansion of the role and function of existing agencies and authorities (such as the General Accounting Office) within or outside of Congress to perform such analyses and evaluations;"

- "improved staffing of standing committees...through the addition of review specialists and persons scientifically trained and experienced in the area of program audit and evaluations;" and

- "the establishment of a central staff or office as an integral part of the Congress charged with performing program audits, evaluations, and analyses under the supervision of a legislative auditor and under the general supervisory control of a joint committee of the Congress..."^{65/}

^{64/} Title VII - Program Review and Evaluation, Conference Report, H.R. 7130, Congressional Budget and Impoundment Control Act, passed Senate June 21, 1974. P.L. 93-344.

^{65/} Sec. 1 (b), H. Con. Res. 21, January 3, 1973.

F. Issues Surrounding the Role of the General Accounting Office in Providing the Congress with Improved Program Evaluation Capability

The General Accounting Office's role in providing the Congress with program evaluations independent of the executive branch has received considerable attention with the last two years. Several issues are involved: (1) Are the agency's auditing, accounting, and program evaluation related activities sufficient to satisfy legislative requirements for program evaluation and evaluation of the utility of different program intervention strategies in meeting legislative goals? (2) Does the agency have sufficient authority, staffing and appropriate administrative arrangements to conduct program evaluations? (3) How can the GAO and the Congress work more closely in developing program objectives and legislative language which meets requirements for the conduct of program evaluation?

1. Basis of GAO's Statutory Authority for Program Evaluation. Comptroller General Elmer B. Staats, noted the basis of the agency's statutory authority for the conduct of program evaluation in a recent article: "While the GAO has always construed the Budget and Accounting Act and the Legislative Reorganization Act of 1946 to include this authority, the Legislative Reorganization Act of 1970 made it quite explicit. This act, in brief, directed that GAO, either on its own initiative or at the request of committees of the Congress, make studies of the costs and benefits of existing programs."^{66/} He continued that the agency has actively carried out its responsibilities under these acts:

For the past 5 years, GAO has given high priority to the evaluation of Federal programs to the point where approximately 30 percent of its professional staff of 3,250 is now engaged in evaluations and studies with this objective.^{67/}

2. Evolution of GAO's Program Evaluation Activities. Issues have been raised regarding the nature of GAO's evaluation activities. For instance, Senator Roth,

^{66/} Staats, Elmer B. Improving Congressional Control Over the Federal Budget. The GAO Review, Summer 1973: 11.

^{67/} Improving Congressional Control Over the Federal Budget, op. cit.: p.11.

commenting upon responses received in his staff survey of program evaluation practices noted:

A rather small portion of the executive departments, and an even smaller portion of the independent agencies indicated in response to our inquiry that GAO was actively or regularly involved in evaluating the substantive accomplishments of their programs. They also stated that the Comptroller General's interest in their programs was quite often of a fiscal-procedural nature.^{68/}

Continuing, he recommended that the GAO's program evaluation responsibilities and those of the other congressional information support agencies, including the Library of Congress and congressional committee staffs, should be expanded to include more program evaluation impact analysis:

It is clear that GAO has plenty of work to do and does much of it effectively. However, Congress needs to have more independent evaluation of the impact of Federal governmental activity—by GAO, the Library of Congress, its own committees, or perhaps by some other body. The Legislative Reorganization Act of 1970 clearly assigns to the Comptroller General and the Library of Congress additional responsibility to perform substantive evaluations.^{69/}

In answering questions posed by critics of GAO's program evaluation efforts, the Comptroller General admits that "GAO still has much to learn," but that "overall good progress is being made." "Evaluation of government program results," he continues, "is an art about which all of us have much to learn." The agency faces two major difficulties in conducting evaluations, "particularly in the social action areas:"

Not the least of these are (1) the lack of clearly, specifically stated program goals and objectives and (2) the lack of reliable data on performance and effects or results of program operations.^{70/}

In another article Assistant Comptroller General Ellsworth H. R. Morse reported that the agency is attempting to improve training and administrative procedures to assist GAO in meeting its program performance and evaluation responsibilities:

^{68/} Public Program Analysis and Evaluation for the Purposes of the Executive and the Congress, op. cit.: S9026.

^{69/} Idem.

^{70/} Improving Congressional Control Over the Federal Budget, op. cit.: 11-12.

1. It is learning much by doing....
 2. It is building an interdisciplinary staff. From a professional staff largely of persons with backgrounds in accounting and conventional auditing, GAO is shifting to a staff made up of many disciplines--including engineers, economists, mathematicians, statisticians, actuaries, systems analysts, persons with academic or experience backgrounds in business or public administration, and a sprinkling of experts in other fields. Over 20 percent of the professional staff is composed of persons trained in disciplines other than accounting and this proportion is growing.
 3. It makes extensive use of consultants in various fields to provide expert advice and assistance.
 4. It contracts work out on occasion.
 5. It conducts extensive training programs and holds special seminars on program evaluation in different fields. ...Seminars have been held on health, welfare, education, and law enforcement assistance.
 6. It takes advantage of on-going or completed analytical and evaluation work of other Federal agencies and outside organizations, such as the Urban Institute and the Brookings Institution. It is developing a guide to program evaluation studies....
 7. It has assembled an advisory panel of about 20 systems and program analysis specialists who are expert in numerous fields of government operations and evaluation.^{71/}
3. Illustrations Provided by the GAO of Its Program Evaluation Activities.

Assistant Comptroller General Morse has identified examples of different types of program evaluations the agency has performed.^{72/} One of these is inserted below to illustrate the agency's conception of its activities in this area:

Improvement of definition of a programs goals and objectives:

Title I of the Elementary and Secondary Education Act of 1965 authorized Federal financial assistance to local educational agencies for programs to help meet special educational needs of deprived children in areas with high concentrations of children from low-income families. The Federal Office of Education required local agencies to establish measurable objectives and related procedures to evaluate the success of their projects. GAO's audit of this program in Illinois showed weak performance on this score. Objectives stated were

^{71/} Morse, Ellsworth H. Jr. The Auditor Takes a Program Evaluation. The Federal Accountant, v. XXII, no. 2, June 1973: 6-7. Emphasis in original.

^{72/} Most GAO reports are public information. In addition short abstracts are published in annual GAO reports.

generally vague (e.g., build a varied vocabulary) and were not stated in measurable terms by types and degree of changes sought (e.g. rate of increase in vocabulary).^{73/}

4. Other GAO Program Evaluation Activities. GAO has also published two reports which deal more directly with improving the process of conducting program evaluations. These, described in detail in other sections of this report, consist of (1) a list of legislative requirements for program evaluation, including steps the Congress needs to take to state program objectives in a form compatible with evaluation capabilities and (2) an inventory of agency information systems which include program evaluation reports to support better dissemination of program evaluation reports to other researchers and to other agency administrators. The agency is now preparing a guidebook of recommended standards for program evaluation. It will be published in 1975.

5. Activities and Legislation in Support of Giving GAO More Responsibility for Program Evaluation. The General Accounting Office's program evaluation activities seem to have increased considerably since passage of the Legislative Reorganization Act and the Budget and Impoundment Control Act. Several factors might be considered in assessing these expanded responsibilities and functions. Probably the most important are:

-psychological, organizational, and methodological factors which surround the evident transition in the agency from considering evaluations as primarily easy-to-measure auditing activities to conceiving of evaluations as harder-to-measure achievement, impact and performance activities;

-steps the agency is taking to enlarge its staff to perform program evaluations and to enlarge its staff to include more social scientists to provide the interdisciplinary mix required for program evaluation; and

^{73/} Morse, op. cit., p. 8.

-the time lag between establishment of new legislative authority and the establishment of effective operational procedures to meet new requirements.

As noted above, under the terms of the Congressional Budget and Impoundment Act, passed June 21, 1974, GAO was given additional authority to conduct program evaluation and to engage in other activities in support of expanding legislative branch program evaluation capabilities. The agency is authorized to establish an Office of Program Review and Evaluation. Among its functions are the following:

- to review and evaluate Government programs and activities, at the initiative of the Comptroller General or at the request of either House or Committees;
- to assist committees in developing statements of legislative objectives and goals and methods for assessing and reporting actual program performance in relation to such legislative objectives and goals; and
- to recommend methods of assessment, information which should be reported, responsibility for reporting, frequency of reports, and feasibility of pilot testing; and
- to review the program evaluation reports and activities of Federal agencies.^{74/}

^{74/} Congressional Budget and Impoundment Control Act, Title VII, op. cit.

PART IV. THE PROCUREMENT AND CONDUCT OF PROGRAM EVALUATION RESEARCH

In the most general sense, program evaluation is a form of applied social science research which uses the interdisciplinary resources and techniques of economics, psychology, sociology, and political science. Some program evaluation research is performed in-house by social scientists on an agency's staff. These evaluations generally are of two types: (1) auditing activities to assess the allocation and use of program funds; and (2) critical syntheses of the findings of the evaluation studies performed for the agency to generate recommendations for improving the administration and operation of a program or to determine whether to modify or terminate the programs.

Evaluations of program performance or of the effectiveness and impacts of a particular program intervention strategy are performed generally on an extramural basis by social scientists or interdisciplinary program evaluation research teams funded by the agency. The Office of Management and Budget has defined four general types of evaluation research, delineated on the basis of their purpose and predominant methodology. These generally correspond to the categories used by Wholey, et. al. in their landmark study on this topic.^{75/} The types are substantive impact evaluations; relative effectiveness evaluations; process or management evaluations; and project evaluations. The OMB definitions of these types of evaluation are given below:

Substantive impact evaluations attempt to measure the impact which federal programs have upon their stated objectives. This type of evaluation seeks to determine what the program accomplishes, how these accomplishments compare to their intended purposes, and their costs. The purpose of such evaluations is primarily to provide information for use in major policy formulation.

Relative effectiveness evaluations seek to compare the effectiveness of two or more major program strategies or approaches in attaining ultimate objectives within a national program. These studies are designed to help policy

^{75/} Wholey, Joseph S., et al. Federal Evaluation Policy: Analyzing the Effects of Public Programs. Washington, D.C., The Urban Institute [1970] chap. 2.

officials and program managers select the most effective mix of services to maximize programs' total impact, such as the mix of skill training, remedial education, and job search assistance in a manpower program. However, these studies do not necessarily measure the impact of the total program in absolute terms on its objectives.

Process or management evaluations are designed to measure the operating efficiency of national programs. They are intended principally to help program managers achieve the most efficient deployment of available resources, rather than help policy officials arrive at major decisions affecting the scope and focus of the national programs.

Project evaluations are directed to individual, locally based projects which are components of a national program, regarding the impact or efficiency of the total national program. Project evaluations may entail any of the three preceding types (substantive impact, relative effectiveness, or process evaluation) as well as project rating comparing the effectiveness of one or more individual projects against others.^{76/}

Four types of research performer are used to perform extramural evaluation research activities. The first two are: (1) social scientists supported by grant or contract to a non-profit making institution; and (2) social scientists at federally funded research and development centers, which are R and D performing organizations usually created and exclusively or substantially funded by one or more Federal agencies to provide the agency with a centralized R and D capacity needed to develop and evaluate its programs. Typical of these are the Urban Institute and the Institute for Poverty Research at the University of Wisconsin, whose major functions were to supply expert information and advice on urban and poverty programs, respectively. A third type of performer consists of university-based social scientists. The fourth type of performer is social scientists or interdisciplinary teams (including operations researchers and systems analysts, etc.) employed by profit-making research or consulting firms working under contracts or grants awarded on a competitive or non-competitive basis by an agency.

A. Issues Relating to the Funding of Program Evaluation Research

1. The Lack of Appropriate Information. The executive branch does not maintain

^{76/} Lewis and Zarb, op. cit.: 309.

an inventory describing precisely how much money has been obligated for program evaluation research. Frequently there are discrepancies in the data systems used. For instance, Laurence E. Lynn, Jr., former Assistant Secretary for Planning and Evaluation, Department of Health, Education and Welfare, noted in 1972 that "The Department has approximately 125 full-time professional evaluators and 40 to 50 million dollars available for evaluation functions."^{77/} According to a report cited by Albert D. Biderman and Laure M. Sharp, the costs for total HEW evaluation research amounted to \$29.6 million in 1971.^{78/}

Annual Budget Appendix volumes sometimes report an entry for program evaluation. However, this item is not reported for all social programs listed in the Catalog of Federal Domestic Assistance. For instance, the Fiscal Year 1974 budget appendix illustrates these discrepancies for three agencies which have large program evaluation research obligations: (1) Health Services and Mental Health Administration, lists an item for research, but not program evaluation; (p. 379); (2) Law Enforcement Assistance Administration lists have no entries for either research or program evaluation (p. 627); (3) Manpower Administration, Department of Labor, lists an item entitled Planning, evaluation, and research, (p. 631).

Biderman and Sharp attribute the dearth of information on the funding of program evaluation research to the fact that existing scientific research reporting categories, including those used by NSF, do not include the classification "program evaluation research"; and there are no well-established mechanisms within agencies to obtain data on the funding of such research:

All available statistical data on the organization, financing, personnel and functions of research utilize categories that have no close correspondence

^{77/} Lynn, Laurence E. Notes from HEW. Evaluation, Fall, 1972: 26.

^{78/} Citing: James G. Abert. Cost and Cost-Effectiveness. Paper delivered at International Federation of Operational Research Societies International Cost-Effectiveness Conference, Washington, D.C., April 12-15, 1971. (Unpublished.) From: Biderman, Albert D. and Laure M. Sharp. The Competitive Evaluation Research Industry. Washington, D.C., Bureau of Social Science Research, Inc. [1972] p.21.

with the world of SPER.* The primary sources of data on American research activities are oriented to the academic science system or to medical and technological research and development. SPER falls neatly into none of these. That most used source for charting trends in American research activity, the National Science Foundation's series, Federal Funds for Science, is of scant use for charting trends in SPER. Even if cross-tabulations of data of this series were available by field of science, funding agency, and research performer, which they are not, we would have no way of knowing which table cells would include SPER.

While, by definition, all SPER is social research, not all of it is reported as social science research and, since it is often supported from operational program funds, an unknown proportion of the total is not reported as science or research at all. Further, allocations to evaluation work are frequently made by the executors of federally-financed action programs or projects--state, local or private agencies which may be layers removed from the point at which book-keeping entries are made that are reflected in NSF's Federal Funds for Science.

Although there is considerable literature arising from interests in government contracting, including studies of R & D procurements (Report to the President. . . , 1962), studies have not focused on what until recently has been that relatively tiny part of all procurement which involves social scientific or related work. Other than those dealing with contracting for developments of defense weapon systems, the major studies have dealt primarily with the National Aeronautics and Space Administration and the Atomic Energy Commission (Senate Committee on Labor and Public Welfare, 1965; Orleans, 1967). When government financing of social scientific work has been the object of attention of studies (and data series), the preoccupations have been almost exclusively with the forms of financing which matter most for institutions of higher education and, secondarily, for the academically-oriented institutes. The types of organizations which perform most of the work of SPER have received relatively little consideration.

In the recent large survey to examine the state and future of the Social and Behavioral Sciences (Behavioral and Social Sciences Study Committee of the National Academy of Sciences and the Social Science Research Council, 1969), for example, a comprehensive questionnaire survey was carried out of the research finances of academic departments, professional schools and university institutes and laboratories, but only a small pilot study in one city--scarcely analysed or reported--was performed of social scientists in all other institutions.⁷⁹

The Survey Committee contracted with the Bureau of Social Science Research in Washington, D.C. for a questionnaire survey of university activities in the behavioral and social sciences which were reported in three chapters of this report (pp. 155-210). A small pilot survey was also undertaken by the committee in the Boston, Massachusetts area to study the employment of behavioral and social scientists in two disciplines (economics and psychology) outside the university and government; this was reported in two pages of the report (pp. 216-218).] Such studies as have been done of social scientists in non-academic employment (Radon, 1970; Marcson, 1966) have involved small samples including many persons working in non-research roles, so that it would be impossible to segregate out for analysis those active in particular research pursuits, such as SPER.^{79/}

* Social Program Evaluation Research

^{79/} The Report on Survey of Federal Program Evaluation Practices, conducted by the staff of Senator William V. Roth, Jr., May 1972, illustrates the difficulties in trying to obtain comparable figures from Federal agencies even when a well-formulated questionnaire is used. Quotation from Biderman and Sharp, op.cit. p.16-

B. Difficulties of Obtaining Information on Types of Performers of Program Evaluation Research

One of the most important implications of the absence of information is that it is impossible to provide precise data on the characteristics of performer, that is in-house, or extramural (either profit-making or non-profit-making research organization). A number of attempts have been made; they have involved in-depth surveys or collation of scattered in-house report forms.

Most of these studies indicate that the bulk of the largest awards for evaluation are given on a contract, rather than a grant basis, to profit-making or non-profit-making research organizations rather than to university based social scientists.^{80/} This finding should be weighed against studies which indicate that evaluation studies are performed best by social scientists at universities funded on grant, rather than a contract basis.^{81/}

For example, Abert reports that about 45 percent of the funds HEW awarded in 1971 for program evaluation went to profit-making research organizations. The recipients of other funds are broken down as follows: not-for profit firms, 29 percent; universities, 21 percent, state and Federal government, 4 percent; and individual consultants, 1 percent.^{82/}

In the Spring of 1974, the General Accounting Office conducted a telephone survey of federally funded program evaluation on behalf of the Joint Committee on Congressional Operations. The results follow:

^{80/} Biderman and Sharp, op. cit.

^{81/} This issue was discussed for example at the "Dartmouth/OECD Seminar on Social Research and Public Policies," Dartmouth, September 13-15, 1974, especially in a presentation by Howard Freeman, based upon research he, Illene Bernstein and Patricia Riecker are doing for the Russell Sage Foundation. Draft report available from: Illene Bernstein, Department of Sociology, University of Indiana, and Dr. Freeman, director, Institute for the Social Sciences, UCLA. For a short summary of the report, see: Evaluation Revisited, Behavior Today, May 27, 1974: p.148.

^{82/} Abert in Biderman and Sharp, op.cit., p.23.

The response to our survey, which are summarized in the following table were for the most part carefully researched and prepared by the agencies listed. A few of the agencies responded in writing-DOT, Agriculture, OEO, VA, and the Special Action Office for Drug Abuse Prevention. Problems with the definitions of program evaluation were fewer than we expected. We believe the total estimate of \$130 million budgeted for program evaluations in Fiscal Year 1974 is fairly accurate for those agencies listed, given the limitations of a telephone survey.

PRELIMINARY ESTIMATES OF FUNDS BUDGETED FOR PROGRAM EVALUATION, RESPONDING AGENCIES, FISCAL YEAR 1974

(In thousands of dollars)

Agency	Amounts budgeted for evaluation breakdown of total by-		
	Contract	In-house	Total
Department of Health, Education and Welfare	50,000	10,000	60,000
Department of Transportation	7,095	18,875	25,970
Environmental Protection Agency	5,756	6,667	12,423
Department of the Interior	3,800	7,584	11,384
Department of Labor	8,900	567	9,467
Department of Commerce	1,031	4,520	5,551
Law Enforcement Assistance Administration	4,337	165	4,502
Veterans' Administration	38	3,154	3,192
Department of State (except AID)	13	3,091	3,104
Department of Agriculture	2,000	1,075	3,075
Office of Economic Opportunity	1,550	1,023	2,573
Department of Housing and Urban Development	1,300	327	1,627
Agency for International Development	528	1,098	1,626
Special Action Office of Drug Abuse Prevention	869	653	1,522
Total	87,217	58,799	146,016
Percent of total	59.7	40.3	

Agencies unable to give a complete response in the time allotted for this preliminary telephone survey: Department of Defense, Department of Justice (Except LEAA), Department of Treasury, National Aeronautics and Space Administration.^{83/}

The agency did not give a breakdown by types of contract performer (that is profit or non-profit making). It seems evident that a continuing data system be established to collect information on federally-funded program evaluation and that it also categorize the type of performer.

^{83/} In U.S. Congress, Joint Committee on Congressional Operations. Congressional Research Support and Information Services: A Compendium of Materials. Washington, U.S. Govt. Print. Off. [May 1974] p.290.

C. Specific Issues Relating to Federally Funded Research and Development Centers as Performers of Program Evaluation Research

The Congress has not given in-depth examination to the conduct of social program evaluation research by federally funded research and development centers (FFRDC). It is difficult to generalize about these activities, except to say that some agencies have used them more than others; and that the use of such centers in the social policy area is probably declining.^{84/} Some information on the pros and cons of the use of FFRDCs was provided in a 1971 report by the National Academy of Sciences. A special advisory committee reported to the Academy on the operations and implications of one such center, the Institute of Poverty Research, University of Wisconsin, sponsored by the Office of Economic Opportunity. The findings of this study are specific to this particular center but could characterize the activities of other centers.

According to the Committee, policymakers should consider a number of issues which surround the conduct of program evaluation research of this nature:

- the "danger of overselling the fruits of even well-conducted evaluations and field experiments;"
- the "strong partisan political currents" that interact with the conduct and use of reports prepared by such centers;
- the need for replication of experiments initially conducted by the center;
- a "division of responsibility for the conduct of an evaluation or experiment between two or more teams so that the knowledge of each would serve as a check against the oversight or misjudgments of the others";

^{84/} The Office of Education made extensive use of such centers; they were phased out at the end of the fiscal year 1972. During the period 1971-1973 HEW, exclusive of the Office of Education utilized 17 such centers; during the period 1972-1974, the agency funded only 13 centers. (U.S. National Science Foundation, Federal Funds for Research, Development, and Other Scientific Activities, Washington, U.S. Govt. Print. Off., 1972, p. 46. (NSF 72-317); and U.S. National Science Foundation. Federal funds for Research, Development, and Other Scientific Activities. Washington, U.S. Govt. Print. Off., 1974. pp. 51-52. (NSF 74-300).

- the need for "open access to the data for subsequent recalculation, criticism and evaluation by independent groups as required;"

- the need to time appropriately General Accounting Office and executive agency investigations of a center's activities or research progress and preferably delay such reviews until an evaluation is complete, since "the threat of [untimely] investigations of this sort...might discourage investigators of high caliber from involving themselves in such policy-related work;"

- the need for policy makers at all levels of government to lower expectations about the utility of policy-oriented program evaluation research since "social science research is able to provide only very limited illumination;"^{85/} and

- the need to consider a number of trade-offs between the advantages and disadvantages of research conducted by policy institutes and to provide the institutes with "long-term, relatively unconstrained funding" since such institutes cannot do "first-rate independent research [with] hand-to-mouth financing."^{86/}

The Commission on Government Procurement, created pursuant to P.L. 91-129, to study and recommend to the Congress methods "to promote the economy, efficiency, and effectiveness" of procurement by the executive branch, looked at the procurement of applied social science research and program evaluation services. The report, presented to the Congress in December 1972, recommended that FFRDCs be organized and used to satisfy needs that cannot be satisfied effectively by other organizational re-

^{85/} "While social science can help in the design and improvement of programs, good policy-making requires art, judgment, and weighing of conflicting values. It can never be reduced to [social] engineering in the familiar sense of the term." (Policy and Program Research In a University Setting: A Case Study. Report of the Advisory Committee for Assessment of University Based Institutes for Research on Poverty, Division of Behavioral Sciences, National Research Council. [Washington, D.C., National Academy of Sciences, 1971] p. 54.)

^{86/} Ibid., pp. 47-55.

sources. Any proposal for a new FFRDC should be reviewed and approved by the agency head and special attention should be given to the method of termination, including ownership of assets, when the need for the FFRDC no longer exists. Existing FFRDCs should be evaluated by the agency head periodically (perhaps every three years) for continued needs.^{87/}

D. Issues Relating to Funding Mechanisms (Grants and Contracts) in Program Evaluation Research

1. Issues Relating to Profit- Versus Non-Profit-Making Performers. Albert D. Biderman and Laure M. Sharp, Bureau of Social Science Research, are conducting an in-depth study on the performance of program evaluation research by both profit- and not-for-profit research institutions. Their interim report, dealing primarily with profit-making research organizations was printed in 1972. Preliminary findings raise a number of issues resulting from the conduct of social program evaluation research performed by profit-making organizations. Generally they observe that competitive bidding for program evaluation research through the request for proposal (RFP) system is wasteful of resources and deleterious to sound research since a large number of organizations go through the expensive process of writing unsuccessful proposals. They gave an example, for instance, of a \$4 million procurement for which research organizations spent \$1.3 million in writing proposals.^{88/}

They also cited the need to assess the disadvantages of the ways in which the RFP system tends to overly prespecify tasks or research methodology that a contractor should follow.^{89/} On this same point the National Advisory Council on Education

^{87/} Summary of the Report of the Commission on Government Procurement. (Not the Official Report.) Washington, U.S. Govt. Print. Off., December 1972. p. 46. For a detailed description of the Commission's findings on performers of research and development, see Chapter 3, pp. 13-23 in Report of the Commission on Government Procurement, Vol. 2 (Washington, U.S. Govt. Print. Off., 1973).

^{88/} Biderman and Sharp, op. cit., pp. 38-39.

^{89/} Ibid., passim.

Professions Development reports:

Within limited evaluation funding on the whole, funds for contract research are so overbudgeted and in-house staff positions and funds are so underbudgeted, that good, practical conceptualization and monitoring of evaluation research is thwarted. In some cases even requests-for-proposals must be developed by outside contractors before proposals for research can be called for. ...Requests-for-proposals have often attempted to specify the evaluation research design rather than the questions of or issues to be addressed, or the final character of the research evidence.^{90/}

Other issues raised by Biderman and Sharp and the Advisory Council are: the lack of time agencies give to writing RFP's since many put out calls for contract program evaluation research only a few weeks before the fiscal year ends;^{91/} monopolization of contracting by a few profit-making organizations; overutilization of systems engineers and systems analysts to perform program evaluation, often using imprecise social science tools and techniques; administrative deficiencies in profit-making organizations; the economic efficiency and quality of the product obtained from the use of sole source contracts vs. the competitive bid proposal; and cost overruns in social R and D activities performed by not-for-profit organizations.^{92/}

E. Legislative Requirements to Mandate a Certain Percentage of Funding for Program Evaluation

Another funding issue, which seems to warrant resolution, is that of legislatively mandating how much an agency should spend on program evaluation. There are two prevailing points of view on this issue. One is to continue the practice of having the Congress authorize a fixed percentage of appropriations of a program for evaluation purposes. A number of suggestions have been made about what percentage of funds should be allocated for this purpose. In an interview given shortly before

^{90/} National Advisory Council on Education Professions Development. Search for Success: Toward Policy on Educational Evaluation. A report of the..., June 1974. Washington, D.C., National Advisory Council on Education Professions Development, op. cit. p. 20.

^{91/} Ibid., p. 20.

^{92/} Biderman and Sharp, op. cit., passim.

he left office as Secretary of Health, Education and Welfare, Elliot Richardson suggested that an allocation of .5 percent of appropriations for a program would be adequate.^{93/} In their study of Federal Evaluation Policy, Wholey et. al., suggested that the fixed percentage should vary depending on the substance of the program:

Funding levels. A reasonable evaluation budget..., this study suggests is likely to range from 0.5 percent to 2 percent of the total program budget. [For programs which do not already include a fixed percentage for evaluation] it would appear reasonable ... to allocate up to 2 percent of manpower program funds, up to 0.5 percent of vocational educational funds, and up to 0.5 percent of Title I elementary and secondary education funds to program evaluation.^{94/}

In his interim report to congressional committee chairman, on "Legislative References to Evaluation, 1967-1972," August 11, 1972, the Comptroller General expressed another point of view -- that legislation should not specify what percent of the appropriations should be devoted to evaluation. His statement follows:

...[We] question whether a legislative requirement that a fixed percentage of program funds or a fixed dollar amount to be spent for program evaluation is wise, particularly where the agency may not be in a position to spend the money effectively. Rigid personnel regulation imposed by the executive branch can be a contributing factor; scarcity of trained personnel has undoubtedly been another factor. [In our studies of evaluation for the Congress]...we will give special attention to this [issue] in those cases where Congress has enacted explicit requirements for program evaluation.^{95/}

F. Alternative Institutions for the Conduct of Program Evaluation Research Under Federal Agency Auspices

Although there are no precise figures on the composition of performers of social program evaluation research for the Government, it seems reasonable to estimate that about half of this research is performed by profit-making research groups or consulting firms, approximately 30 percent by non-profit institutions and 20 percent by

^{93/} Conversational Contact...with Elliot L. Richardson. Evaluation, v. I, no. 1, 1972: 15.

^{94/} Federal Evaluation Policy, cit., p. 82.

^{95/} Letter prefacing GAO Report B-161740, August 11, 1972.

universities.^{96/} As noted throughout this report, many criticisms have been raised about the uneven quality of evaluation research. These deal especially with methodological weaknesses, the tenuous conceptual link between program evaluation research design and findings on the one hand, and understanding of the basic dimensions of human behavior on the other, and insensitivity to the policy and political constraints which impact upon the initiation and implementation of program evaluation reports. These criticisms are levied at both university and non-university performers and also at in-house and extramural performers.^{97/} Program evaluations by non-university researchers are generally timely, but weak in theory, conceptual framework and methodological underpinning. Those performed by university researchers are criticized for being overdue and insensitive to policy implications. There seems to be a consensus in much of the recent literature that the conduct and utility of program evaluation research would be improved if the Government adopted a formal policy to institutionalize a sustained program evaluation capability in research organizations which have

^{96/} Extrapolated from a classification of performers of HEW evaluation research, 1971 (The Competitive Evaluation Research Industry, op. cit.: 23.) For further details on funding and performers see the draft reports of Freeman, and Freeman, Bernstein and Rieker, op. cit.: passim.

^{97/} Evaluation magazine contains several articles on this topic:

"On the question of 'contract' evaluations, Garth Buchanan, Pamela Horst, and John Scanlon vote no--that the costs of unsatisfactory evaluations produced by contract are far more, really, than the high dollar cost of in-house evaluation efforts. In Evaluation's interview with him, Bertram Brown votes yes, because the use of outside evaluators has modified the impact of administrators who want 'positive' funding attached to evaluations. Howard Davis, Carol Weiss, and Lawrence Lynn, Jr. are among others who suggest that administrators' desire for positive findings has been known to lead to positive findings, but Weiss is also quick to point out that, whatever the pluses or minuses, the instability of most evaluation efforts and the turnover of evaluation staffs is bound to lead to the use of consultants in one capacity or another."

(See Evaluation, v. no. 2, 1973, passim.)

a strong social science base.^{98/}

Recommendations for improving the institutionalization of program evaluation research take two forms. One is that the Government should support the creation of research units designed explicitly to provide short and long-range program evaluation research for a particular agency or class of problems. The second is a call for giving more sustained funding to university researchers to enable them to develop a better capability to do basic, but policy-oriented, social science research.

Proposals for creating research units to provide evaluation research for a particular agency or class of problems call for institutions of varying configuration. However they could generally be characterized as large-scale non-profit applied social science organizations, either independent of an agency or attached to a Federal agency like the recently created National Institute of Education, or funded by an agency to provide it with information and analysis, like the Urban Institute, created first to serve the Department of Housing and Urban Development. This section summarizes several of the published arguments for the creation of such units.

Hendrik D. Gideonse, director of program planning and evaluation of the National Center for Educational Research and Development of the U.S. Office of Education, proposed, in 1970, that programs for social services would be improved by the development of national research policies and national research organizations to provide information, analysis, and program evaluation research for these programs. Many of the propositions he set forth for a national research strategy in education contain elements which can be generalized to other areas. His points are summarized below:

-The funds a nation spends for educational research and development should be considered a part of the nation's general expenditures for scientific research

^{98/} For example, see the report of the National Advisory Council on Educational Professions, Search for Success, which recommended steps to improve educational evaluation research. Also described in Evaluation Reforms Urged, Behavior Today, v.I, no.2, August 5, 1974:1.

and development. Thus, educational research funds should be subject to the same review procedures used for other types of scientific R & D. Funds for education should compete with funds allocated for other areas of R & D; and priorities for educational R & D should, therefore, be established on a nationwide basis.

-Educational R & D should be interdisciplinary (incorporating research into educational technology, social change, quality of instruction, goals of learning, etc.); thus, such research should be conducted by interdisciplinary teams in large-scale research organizations.^{99/}

-Two factors now discriminate against the conduct of effective educational research by university based social scientists: the reward structure of academic departments and self-generated professional values.

-Goals and priorities for education should incorporate many diverse values and points of view, not only those held by social scientists or educators; therefore national research strategies should incorporate consideration of these other factors.

Gideonse stressed that development of the then proposed National Institute of Education would meet these criteria since it would coordinate the educational research activities of several agencies. However, he cautioned that the agency would have to take steps to meet another criterion of effective educational research units. It would have to "build the political decision structures which...constitute the necessary enabling condition for success."^{100/}

^{99/} For background information on the complexities of conducting interdisciplinary social science research, see: U.S. Congress. House. Committee on Science and Astronautics. Subcommittee on Science, Research and Development. Interdisciplinary Research--An Exploration of Public Policy Issues. A study prepared by the Science Policy Research Division, Congressional Research Service, Library of Congress, October 30, 1970. Washington, D.C., U.S. Govt. Print. Off. 1970. 106 pp. [Committee Print.]

^{100/} Gideonse, Hendrik D. Policy Framework for Educational Research. Science. December 4, 1970: 1058.

Peter H. Rossi outlined a variant of this position when looking at the need to create non-profit interdisciplinary research units at universities or research units under a university consortia mechanism. He noted that program evaluation research required input from a number of disciplines and used a number of techniques. Such research is expensive and requires economies of scale. Only large research units, "with a director of research, his own staff,...a more or less integrated program of research," and drawing upon a number of disciplines would be able to provide research needed for the future.^{101/} These organizations would utilize the talents of academic researchers and thus would have to overcome a number of current obstacles to large-scale interdisciplinary social science research; and social scientists would have to be given a modicum of autonomy similar to that permitted in academic departments. Teams of social scientists, with appropriate research support personnel would have to be created; and social scientists would have to learn to work within the political and social constraints inherent in applied social research, constraints caused by problem definition, the conduct of research, and the procedures to publish and disseminate research findings.^{102/}

The most recent proposal along these lines comes from the Social Science Research Council's Committee on Experimentation as a Method for Planning and Evaluating Social Intervention. When assessing the capability of existing research organizations to do policy relevant social experiments, the committee concluded: "...these existing organizations possess little or no capability to design experiments and manage a

^{101/} Rossi, Peter H. Observations on the Organization of Social Research. In Evaluating Social Programs: Theory, Practice, and Politics. Edited by P.H. Rossi and Walter Williams. New York, Seminar Press, 1972. pp. 267-87.

^{102/} Ibid., p. 283.

flexible field organization."^{103/} The committee recommended an organizational locus for such studies: the provision by the Federal Government of relatively small institutional grants ranging in size from \$250,000 to \$500,000 per year, to between six and ten promising research institutions or a consortium of institutions to enhance existing capabilities.

Walter Williams, a social scientist with long experience conducting and administering program evaluation for the Government described the rationale for this position and his variation of the recommendation. Williams states that the Federal Government should allocate sustained institutional support for this purpose to social science research organizations independent of universities.^{104/} He reasons that university based social science research organizations are incapable of meeting the criteria for

...carrying out...sound large-scale social policy research and policy analysis [which] seems to require organizations with research credibility, staff continuity, a reasonable degree of control over the allocation of staff resources, a strong policy commitment, and a flexible incentive structure that rewards diverse skills in terms of pay and status.^{105/}

University based social scientists, according to much of the recent literature,

^{103/} Quoted in Williams, Walter, The Role of Social Scientists Outside the Government in the Federal Social Policy Process, a background paper for an Address to be Delivered on August 25, 1974, at the Canadian Sociology and Anthropology Association Meetings in Toronto, Canada. Draft: p.26. The committee's full report is available as: Riecken, Henry W. and Robert F. Bouch, eds. Social Experimentation: A Method for Planning and Evaluating Social Intervention. (Written by a Committee of the Social Science Research Council.) New York, Academic Press [1974] 339 p.

^{104/} Williams, *ibid.*, p. 36. See a larger exposition of this position in his "The Capability of Social Science Organizations for Performing Large-Scale Evaluative Research, in Peter H. Rossi and Walter Williams, eds. Evaluating Social Programs, Seminar Press, 1972. p.306.

^{105/} Williams, The Role of Social Scientists Outside the Government in the Federal Social Policy Process, *op. cit.*, pp. 22-23, emphasis in the original. On this point see: Coleman, James. Methodological Principles Governing Policy Research in the Social Sciences, a paper presented to the American Association for the Advancement of Science, December 29, 1972.

do have a potentially substantial role to play in enhancing the capability of social program evaluation research and policy analysis for the Government. According to Williams, "the universities...probably should have the major role in the more fundamental policy-oriented research...[those]...more basic studies [which] offer the traditional incentives of the past..."^{106/} Carol Weiss, another social scientist with extensive program evaluation experience has elaborated on this position:

There may be greater potential in doing research on the processes that give rise to social problems, the institutional structures that contribute to their origin and persistence, the social arrangements that overwhelm efforts to eradicate them, and the points at which they are vulnerable to social intervention. Pivotal contributions are needed in understanding the dynamics of such processes and in applying the knowledge, theory, and experience that exist to the formulation of policy.^{107/}

^{106/} *Ibid.*, p. 37.

^{107/} Weiss, Where Politics and Evaluation Research Meet, *op. cit.*, pp. 44-45.

PART V. METHODOLOGICAL ISSUES IN
THE CONDUCT OF PROGRAM EVALUATION

A number of issues raised in previous parts relate directly to the general topic of this section. These include, for example: composition of the interdisciplinary team conducting the program evaluation; types of performer; legislative language especially in relation to program objectives, requirements for performing evaluations, legislative statements regarding funding, and the compatibility between these statements and program evaluation methodology; proposals to create program evaluation information dissemination systems; and administrative obstacles to the conduct and use of program evaluation.

Three additional issues will be treated here since they are receiving a good deal of attention in the literature on evaluation. They are the appropriateness of evaluation techniques, especially social experimentation, improved techniques of data collection and cross-fertilization between developments in social indicators research and data required for program evaluation.

Experimental and Non-experimental Research Design

Three basic types of research design are used in conducting social program evaluation: non-experimental, quasi-experimental, and experimental designs. Most federally-supported program evaluations probably use non-experimental research designs, which permit two types of evaluation methodology: before versus after program comparison, and comparisons of planned performance versus actual performance. Non-experimental evaluations are the easiest and cheapest to conduct. However, they usually are not planned in advance, making it difficult to collect appropriate data. One of the major deficiencies of such designs is that they fail to control for many of the observed changes which occur during the program's duration. Thus it is difficult to explain whether the program intervention strategy caused the changes, if any, observed in the participants, or whether the

changes were caused by some other extraneous factor or event which affected the program or its participants.

Quasi-experimental designs include evaluations which project pre-program data versus actual post-program data over a period of time or compare changes in program participants with jurisdictions or populations not served by the program. These designs have procedures to control for or isolate extraneous factors which might influence the outcome of the program intervention and utilize a more sophisticated data base and logical or conceptual underpinning than non-experimental designs. However, they are far less rigorous and explanatory than experimental designs.

Experimental designs, called controlled experimentation or social experimentation, are the most costly and difficult types of research used in program evaluation. They also yield the most reliable and valid results. However, they require mandatory planning before the program is implemented in order to build in procedures to collect appropriate data and to select randomly a control group and test population.

See the attached chart for a short description of evaluations using these three types of research designs.

Scope and Limits of Social Experimentation

As noted throughout this report, criticism of program evaluation and its use in decisionmaking is widespread. For example, Joseph Wholey, a staff member of the Urban Institute, and widely published in the field of evaluation research, has summarized some of these criticisms and their causes:

Valid, reliable evaluation is very hard to perform and can cost a lot of money. Evaluators have real problems in detecting causal connections between inputs and outputs and in doing so in timely enough fashion to be useful to decisionmakers. The structure of a program can have an important influence on the technical feasibility of separating the effects of the program from the effects of other, often more powerful, forces not under control of the program. ...Our reviews have found typical Federal program evaluation studies marked by certain design characteristics which severely restrict their reliability

Non-experimental	Quasi-experimental	Experimental
<p>Before vs. After Program Comparison</p> <p>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.</p>	<p>Comparisons of Planned vs. Actual Performance</p> <p>Time Trend Projection of Pre-Program Data vs. Actual Post-Program Data</p> <p>Compares actual post-program data to estimated data projected from a number of time periods prior to the program.</p>	<p>Controlled Experimentation</p> <p>Comparison with Jurisdictions or Population Segments Not Served by the Program</p> <p>Compares pre-selected similar groups, some of whom are served and some of whom are not (or are served in different ways). The critical aspect is that the comparison groups are pre-assigned before program implementation so that the groups are as similar as possible except for the program treatment.</p>

109/ Taken from: Hatry, Harry P., Richard E. Winnie, and Donald M. Fisk, et. al. Practical Program Evaluation for State and Local Government Officials. Washington, D.C., The Urban Institute [1973] pp.39-40.

and usefulness:

- (a) They have been one-shot, one-time efforts, when we need continuous evaluation of programs.
- (b) They have been carried out in terms of national programs and are very weak on process data.
- (c) They have been small sample studies working with gross averages, when we need studies large enough to allow analysis of the wide variations we know exist in ... and performance among projects within programs.

These studies have often been accompanied by conclusions and recommendations based on unsupportable or unmeasured assumptions and weak, and often confusing data. In these cases, the evaluation results should be ignored by policymakers. Other evaluation studies, while competently conceived, are so severely constrained by time, money, and an inadequate data base that the results at best have only limited significance for policy changes or program improvement. 108/

Many critics of current program evaluation practices hold that this situation would be improved if social programs requiring evaluation were designed and implemented in a manner which permitted evaluation using controlled experiments or social experiments. For instance, the Joint Economic Committee conducted a series of investigations dealing with the administration and evaluation of public welfare programs. The difficulty of adhering to the rigorous criteria required in social experimentation was cited as a principal deficiency of these programs:

Improving The Evaluation Process

The robust expenditures for research and evaluation of training programs (\$179.4 million from fiscal 1962 through 1972) are a disturbing contrast to the anemic set of conclusive and reliable findings. Although some of the data may be necessary management information, much of what is collected as a matter of course by program administrators cannot be used to estimate the impact of training and determine the effectiveness of the program.

Among the most glaring deficiencies are inappropriate control groups and short observation periods. An appropriate control group is essential if the impact of training is to be isolated and distinguished from the influence of other factors. Frequently, studies are undertaken without any control group. When a control group is included in the design, it is almost never established at the time that enrollees enter the training program. Evaluators are called in after training has occurred, and are confronted with the almost impossible task of constructing a control group with the pretraining characteristics and experience of the trainees.

Observation periods rarely last more than a year and usually less. This is too short to determine how long benefits last or whether they are stable,

108/ Wholey, Joseph S. What Can We Actually Get From Program Evaluation? Policy Sciences, v.3, September 1972: 367-9.

increase, or decrease. Evaluators are forced to base their estimates of program effectiveness on uneasy assumptions about the duration of earnings gains. Tracking trainees and a control group over an extended period of time is an expensive proposition but it is probably no more expensive and certainly more useful than much of the data collection and evaluation which has been conducted in the past.

Because the structure of programs and the characteristics of their clientele change over time, a single evaluation, even one with an optimal design, conducted at one point in time does not provide reliable information about program effectiveness. A standardized, ongoing evaluation procedure should be established. ^{110/}

Although there seems to be growing dissatisfaction with non-experimental program evaluation designs and considerable enthusiasm for using experimental designs as a program evaluation strategy, it is apparent that a number of expectations and concepts of evaluation must be modified before this approach can be used successfully. One problem, for instance, is that of political bureaucratic resistances to using evaluation of any sort -- because of potential negative consequences for an agency's or a committee's pet project. ^{111/}

Researchers have given considerable attention to outlining ways to overcome some of the bottlenecks which hamper the utility of social experiments in the policymaking process. Some of their positions are summarized below.

Wholey, for instance, notes the need for decisionmakers to be aware of the requirements of experimentation: continuous evaluation, better data, and the need for larger sample populations; and the need for evaluators to be sensitive to the political and time constraints on using experimentation as an evaluation strategy. He states in part:

There are important tensions between the evaluator and the program

^{110/} U.S. Congress, Joint Economic Committee. Subcommittee on Economy in Government. Economic Analysis and the Efficiency of Government. Report. Washington, U.S. Govt. Print. Off. [February 9, 1970] pp.14-15.

^{111/} See for instance, Donald T. Campbell, "Assessing the Impact of Planned Social Change," Background Paper for the Dartmouth/OECD Seminar on "Social Research and Public Policies," September.13-15, 1974. passim.

official, tensions which arise out of the very notion of experimentation. The criteria for selection of sites, the carefully controlled design of the experiments, and the random assignment of participants (or communities) to treatments are basic to experimental design. The program administrator may not see the utility of such ideas, however. What is so wrong, he may wonder, about calling an existing exemplary program an "experiment"? Or why not choose the people most in need of housing to participate in a housing allowance experiment? The evaluator must woo and win the administrator to the need for preserving the experimental character of the experiment.

Time also presents an enormous problem for the evaluator of experimental programs. As soon as there is sufficient legislative support to fund a series of experiments, there may be enough support to enact such a program nationwide. The concern that legislation will be enacted before the experiments have had time to produce reliable results may lead to pressures for the release of early, less reliable findings. The New Jersey Income Maintenance experiments experienced this pressure. Some early tentative results from the study were released with reluctance and heavy qualifications. If experimentation is to become a major vehicle in policy research, then ways must be found to anticipate and deal with these types of pressures. ^{112/}

Dr. Alice Rivlin, senior fellow, Brookings Institution and chairman of the National Academy of Science's Committee on Federal Program Evaluation, is one of the Nation's foremost authorities on using social experimentation to improve program evaluation and ultimately the delivery of effective social services. Her views are contained in a recent publication: Systematic Thinking for Social Action, Washington, D.C., The Brookings Institution, Washington, D.C., 1971, 150 pages. Rivlin discusses the comparative merits of several programs and evaluation activities of them, notably the New Jersey negative income tax experiment funded by OEC, the follow-through of Project Headstart, and performance contract teaching. Despite her advocacy of using social experimentation as a tool of program evaluation, she outlined major obstacles to its use. These include: changes in objectives of the program as it is administered; the need for objective evaluation of the benefits of a program; problems of the Hawthorne effect (that is, the effect of the researcher on the test population or the participants of the program);

^{112/} Wholey. What Can We Actually Get From Program Evaluation? Op. Cit: 367-369.

generalizability of findings from a selected to a general population; political and ethical objections to experimentation; and problems of conducting experiments and evaluation on programs whose impact might not be immediately noticeable. 113/

In a subsequent analysis, published in 1973, Rivlin cautioned that social experiments are a useful tool for decisionmaking, but that successful social experiments are rare. The success of a social experimentation strategy, according to Rivlin, varies with the type of program intervention strategy employed. Overly enthusiastic advocates of experimentation should be aware that this is not an appropriate evaluation strategy for all types of programs. In detail:

It seems clear that experiments are feasible when the treatment is a simple one, such as a change in tax or a payment schedule and when the outcome of interest is measurable behavior of individuals, such as hours worked or dollars earned. The income maintenance, health insurance, and housing demand experiments are all of this type. They pose plenty of problems, but they certainly can be carried out.

It is still an open question whether or not more complex experiments are feasible -- those in which the treatment is a multi-faceted change, such as the introduction of a whole new education curriculum, or where the outcome of interest is a change in a whole institution such as a school system or a housing market. These kinds of experiments are far more difficult and no one is sure yet that they can yield valid results. 114/

"If great care is not taken to make current experiments as useful and as sensible as possible," she concluded, "there may be a reaction against the whole technique and a potentially useful tool may be taken away." 115/

113/ For a review of other studies treating utilization of social experimentation methodology in social programs, see: Wegner, Robert E. C. Evaluating the Evaluators. Public Administration Review [January/February 1973] pp.80-87. The books treated in addition to Rivlin's are: Kershaw, J. A. Government Against Poverty. Chicago, 1970; Levine, R. A. The Poor Ye Need Not Have With You: Lessons from the War on Poverty. Cambridge, 1970; and Williams, Walter. Social Policy Research and Analysis. New York, 1971.

114/ Rivlin, Alice M. Social Experiments: The Promise and the Problems. Evaluation, v.1, no. 3, 1973: 78. In this same respect Walter Williams has presented a strong case for using a policy research strategy, rather than an evaluation strategy, to serve some evaluation requirements. See his: The Direct Usefulness of Evaluative Research and Policy Analysis in Social Policy Making, an address delivered at a conference on "Evaluative Concerns in the Post-secondary Education of Students from Disadvantaged Backgrounds," Mercy College, Detroit, June 22, 1974.

115/ Ibid.

Legislative Interest in the Ethics of Experimentation. It should be noted that legislative interest in the ethical issues impinging on evaluation and experimentation is indicated by the passage of the National Research Service Award Act, P.L. 93-348, 1974. In addition to other functions, the Act established a Commission for the Protection of Human Subjects of Biomedical and Behavioral Research to advise the Department of Health, Education, and Welfare, by identifying ethical principles and guidelines which should underlie the conduct of biomedical and behavioral research involving human subjects. The Commission also will work with the Department to develop administrative procedures to carry out the guidelines. The act also creates a permanent National Advisory Council for the Protection of Subjects of Biomedical and Behavioral Research, effective, July 1, 1976; and requires the Commission to undertake a special study of the ethical, social and legal implications of advances in biomedical and behavioral research and technology. 116/

The National Institute of Mental Health awarded \$75,000 during the summer of 1974 to the Bureau of Social Science Research to develop guidelines for proposed regulations on social science research required under this law. 117/

The Need for Continuous, Appropriate Impact Data

One of the most frequently voiced complaints made by program evaluation researchers is the absence of continuous, appropriate data required for program evaluation. As noted in part two of this series, much of the reporting information required to comply with administrative requirements for Federal grant programs

116/ Legislative History: House Report 93-224, May 23, 1973 to accompany H.R. 7724; Senate Report 93-381, August 3, 1973, to accompany H.R. 7724; Dates of consideration and passage: House: May 31, 1973, June 28, 1974; Senate: September 11, 1973, June 27, 1974.

117/ Behavior Today, September 18, 1974: 236.

CRS - 66

consists of program output information stated in terms of dollars, numbers of people affected, and the like. While this information is appropriate to administrative categories and auditing, it is not useful for program evaluation since it does not portray harder-to-measure performance or impact characteristics. For instance, a program intervention strategy providing education to train unskilled workers and assist them to find employment typically requires collection of information on the number of persons who obtained jobs after the program or improved scores on achievement tests. These data, which may be collected in the course of normal administration of the program are not necessarily a good indication of the program's success and failure or of impacts on the persons involved. More intangible factors and information may have to be considered, such as psychological and sociological changes in program participants which may change outlooks and later pay off in job-hunting and employment opportunities available in the area.

Although generation of data appropriate to measuring a program's performance or impacts is to some extent the responsibility of the program evaluator, many administrators and evaluation researchers hold that procedures should be built into a program to provide for continuous collection of the kind of data required for evaluation. They also note that these data requirements should be considered by the Congress and incorporated into authorizing legislation before the program is implemented.

Researchers at the Urban Institute treated this issue in a report prepared for the Joint Economic Committee. Excerpts are reproduced below. Manpower training programs were the immediate issue:

Manpower Training Information Congress Needs and Should Require

The information about the cost-effectiveness of training programs presently available has been found wanting in consistency and usefulness. What steps can Congress take to assure improvement in the future? To answer this, it is first necessary to consider the several facets of manpower

training programs over which Congress may exercise its review powers. These include:

- * The distribution of training services and benefits among target groups to insure that national distributional goals are met.
- * The administration of the national programs to determine whether the programs are improving over time
- * The questions of costs versus benefits to determine whether the programs are worthwhile investments from various viewpoints.
- * The use of manpower training programs as economic tools in dealing with unemployment and inflation.

The most valuable information for the purposes of the Congress may well be data that permit comparisons to be made from year to year to detect trends in identically calculated measures of the programs. To improve the accuracy and consistency of such data, the Congress should require that information be gathered in a format similar to that shown in a recent Joint Economic Committee publication. ^{118/} The format for distributional data shown there should be adequate and could be provided currently by the Department of Labor. It would modify the cost and benefit data to include for each program: ^{119/}

- * Average cost per enrollee.
- * Number of enrollees.
- * Number of completers (by categories in those programs that have multiple categories of completers).
- * Number of employed completers.
- * Average workweek of an employed completer.
- * Increase in income of employed completer.

All benefit data should be accompanied by an explanation of the method of determination and a specification of the groups of participants (completer, employed completer, dropout, etc.) to which it is applicable.

...We recommend that unsupported assumptions on the value of work performed, social costs, private costs, or projected benefits not be used in the benefit/cost calculations. However, we do recommend that attempts be made to measure these various elements.

Each year, the Congress should also request the Department of Labor to examine raw cost and effectiveness data accumulated at the appropriate administrative level for all projects within the programs and use these for determining the causes of variations. ^{120/} These data could then serve as the

^{118/} Economic Analysis and the Efficiency of Government. Op.cit., appendix 3: Illustrative Manpower Overview Data.

^{119/} For the NYC program, measures of its goals of keeping participants in school would also be needed.

^{120/} For programs like CEP and WIN, the appropriate administrative unit would be the individual CEP or WIN project. In the case of MDTA, data might be more properly accumulated at the State level. As decentralization proceeds, the relevant administrative unit might become the local prime sponsor of each comprehensive manpower program. This should improve the capability for review both in DOL and in the Congress.

basis for annual discussions between Congress and the executive branch on the improvement of performance within each program.

If the information provided is to be continuous and to grow in accuracy and consistency, several features must be mandated through legislation. All manpower legislation should include provision for:

- * One standard, comprehensive system of required reporting from the local level.
- * One-year follow-up on the wage rate, income, and job stability of participants.
- * Systematic processing at several administrative levels to provide plans vs. performance information and relative effectiveness information on local projects.
- * Reduction of this information to the formats desired by the Congress.

...To enable more knowledgeable use of manpower training programs as economic tools, there is great need for validated quantitative models that describe the impact of trainees from manpower programs on local labor markets. It is time to stop talking about what effects trainees might have on local markets and measure what these effects are in practice, at least on a small scale. Out of such testing could come more definitive advice for both the Department of Labor and the Congress about the economic effects of training programs. We recommend that Congress direct and earmark sources for data collection projects focusing on measurement and analysis of the effects of training programs on local labor markets under different conditions. ^{121/}

Legislative efforts to improve the collection and use of program evaluation related information are reflected in Title VIII of P.L. 93-344, the Congressional Budget and Impoundment Control Act. Under this Title, the Secretary of the Treasury and the Director of the Office of Management and Budget are to cooperate with the Comptroller General and the Congressional Budget Office in developing and using budget information systems which include "program-related information." Such budget information is to be furnished to congressional committees. The General Accounting Office is also authorized to develop, establish, and maintain an up-to-date inventory and director of sources and information systems containing fiscal, budgetary, and program-related data and information and a brief

^{121/} Nay, Joe N., Scanlon, John W. and Wholey, Joseph S. Benefits and Costs of Manpower Training Programs: A Synthesis of Previous Studies With Reservations and Recommendations. In U.S. Congress. Joint Economic Committee. Benefit-Cost Analyses of Federal Programs. A Compendium of Papers. January 2, 1973. Washington, U.S. Govt. Print. Off., 1973: pp.259-61.

description of their contents.

Cross-fertilization Between Advances in Social Indicators Research and Improvements of Data Required for Program Evaluation

A related issue is that of the relationship between social indicators and requirements for improved data for program evaluation. Criteria for the development of refined social data for program evaluation purposes closely parallel the objectives and expectations of social indicators research. A short review of the relationship is given below. In addition, major obstacles to the development of social indicators are noted. It should be further emphasized that many of the difficulties of developing social indicators also characterize the refinement of program evaluation data.

According to social indicators researchers, social indicators would overcome the inadequacies of currently collected social and economic data and would provide for the measurement or quantification of qualitative or "soft" social factors. ^{122/} Current social indicator research activities are designed to develop indicators which would reflect the following properties:

- (1) They would give a more realistic representation, or a better proxy than do current data series, of the phenomena or causes and effects which they purport to represent. For instance it has been said that the expansion of food programs for the needy does not adequately indicate whether hunger and malnutrition are being eradicated. Alternative data would describe the effects of such programs on such factors as improvements in health among populations served by food distribution and funding

^{122/} For a review of current social indicators activities and an enumeration of policy issues relating to their development, see: G.J. Knezo. Social Indicators: A Review of Research and Policy Issues (Including Annotated Bibliography). Science Policy Research Division, Congressional Research Service, Library of Congress, Multilith 73-39 SP., February 9, 1973, 127 pp.

programs. ^{123/}

- (2) They would differ from currently collected social statistics because they would measure disaggregated and often subjective factors, such as the amount of happiness or health in different racial or ethnic groupings scattered throughout the country.
- (3) They would be normative and indicate changes in the quality of life. The fact that attention is given to the collection of data about the purity or impurity of the air, city by city, or region by region, can be interpreted as an indication that the society collecting this information recognized purity of air as a social goal.
- (4) They would be comparable, longitudinal, and additive. For instance the value of job satisfaction to a particular individual would be measured or weighed against the satisfaction or pleasure he derives from having ready access to recreational facilities. Time series data, or data collected over a number of years, would indicate whether one or another type of satisfaction is more important at some particular time such as in youth, middle age, or old age.

The objective of developing social indicators, with these properties, is to

^{123/} According to Dr. R. Mendelsohn, Director of Community Pediatrics, University of Illinois College of Medicine, there are several widely used, but misleading, indicators of whether hunger and malnutrition are being conquered. These include: expansion of plans for food stamp distribution programs; increased appropriations for school lunch programs, and increased funding of anti-poverty programs. As an alternative Dr. Mendelsohn contends that we should collect and measure the possible short and long-term effects of such programs. These new data series would include information on: decrease in infant mortality rates between families who are supported by these programs and those that are not; differences in height between poverty program children and those not on poverty programs; and prevalence of mental retardation between those on and off poverty programs. (Testimony of Dr. Mendelsohn, Select Committee on Nutrition and Human Needs. In Malnutrition and its Effects. Statement of the Honorable Charles Percy on the Floor of the Senate. Congressional Record. April 30, 1970: p.6414.)

provide program administrators, and policymakers with more accurate and valid social data to draw upon in evaluating the progress of social programs and to prepare annual balance sheets describing the social state of the Nation and alternative solutions to major social problems.

An indication of the importance of this issue to the executive branch is indicated by publication of the Office of Management and Budget's first social indicator report, Social Indicators, 1973. ^{124/}

It should be pointed out, however, that there is a continuing debate within the social science communities about the immediate policy utility of social indicators research. Some social indicators researchers hold that, imperfect as they are, social indicators should be used in policy formation. ^{125/} Others suggest that social indicators are not now of immediate utility to policymaking, but that their utility would be enhanced if more research were conducted on subjective dimensions or people's perceptions about the quality of life. ^{126/} Some recommend that the utility of social indicators in policymaking will be hastened if greater efforts are made to integrate social indicators research

^{124/} U.S. Office of Management and Budget, Statistical Policy Division (with the Social and Economic Statistics Administration, U.S. Department of Commerce). Social Indicators, 1973: Selected Statistics on Social Conditions and Trends in the United States. Washington, U.S. Govt. Print. Office, 1974. 258 pp.

^{125/} For example: Taeuber, Richard C. Social Indicators 1973: Some Policy Considerations. Paper presented at the August 1974 meetings of the American Statistical Association, 12 pp.; and Mondale, Walter F. Social Accounting: Evaluation and the Future of Human Services. Evaluation, v.1, 1972: 30.

^{126/} For example, see the reviews in: Parke, Robert and Sheldon, Eleanor Bernert. Social Indicators One Year Later: An Overview. A paper prepared for presentation at the meetings of the Second Annual Social Indicators Conference, sponsored by the Public Policy and Issues Division, American Marketing Association, February 21, 22, 23, 1973. Washington, D.C., mimeo, 14 pp.; and Sheldon, Eleanor Bernert and Parke, Robert. Social Indicators: Social Science Researchers Are Developing Concepts and Measures of Changes in Society. Mimeo, forthcoming in Science, 34 pp.

and federally sponsored program evaluation activities. ^{127/} Some researchers have, in fact, faulted the OMB's first effort at compiling an annual social report, its study Social Indicators, 1973, because the data presented were not interpreted or evaluated in a manner which gave policymakers any indication of how present social conditions or changes in social conditions overtime related to the problems they face. ^{128/}

Some recognition of the need to develop social indicators for policy-making purposes is reflected in "The Full Opportunity and Social Accounting Act," which was approved in the Senate in both the 91st and 92nd Congresses, and was reported and tabled in the 93rd Congress. ^{129/} The proposal provided for creation of a Council of Social Advisors in the Executive Office of the President. The council "would be responsible," according to Senator Walter F. Mondale, the bill's principal sponsor, "for monitoring, on an on-going basis, specific actual conditions in the country which affect the 'social opportunity' of our people." "Developing a system of social indicators," the Senator continues, "would be a principal task of the council." Utilizing data generated by the Council of Social Advisors, the President would be required to report annually "on the Nation's social status." The following areas would receive priority: "education, health, alienation, political participation, personal security, and social mobility." ^{130/}

^{127/} Schneider, Jerrold E. Making Government Work: Political Versus Technical Obstacles to Social Accounting. American Behavioral Scientist, v.17, n.4, March/April 1974: 585-607; and Mondale, op. cit.

^{128/} Tauber, op. cit.: 6-7.

^{129/} S. 5, The Full Opportunity and National Goals and Priorities Act, statement on the floor of the Senate by Sen. Walter F. Mondale. Congressional Record. January 4, 1973: pp. S52-56. The bill, S. 5, was reported by the Committee on Labor and Public Welfare in July 1973 (Sen. Report No. 93-324), and saw floor action which deleted original Title II of the proposal which proposes to create a Congressional Office of Goals and Priorities Analysis.

^{130/} Mondale, Walter F. Social Advisors, Social Accounting and the Presidency. Law and Contemporary Problems [Duke University School of Law, Summer 1970]: 497.

Such a system is necessary, Senator Mondale says, to give "societal knowledge visibility of the sort that only presidential involvement can generate," and to provide Congress and the President with better data on which to make more informed decisions.

Recent literature and research reports on social indicators have identified a number of other obstacles to the development of social indicators and improved program evaluation data. These are described next.

Understanding of cause and effect relationships underpinning changes in society is another criterion which must be met before social indicators can become useful for policymaking. However, according to Sheldon and Land, "...the state of the art with respect to the development of explanatory relationships [in the social sciences] is especially deficient."^{131/}

Another of these problems is the issue of invasion of privacy. Longitudinal or time-series data, in some cases, requires annual interviews of the same persons on particularly sensitive questions. It is entirely conceivable that individuals would resist questions by Government officials regarding happiness, marital problems, and the like. Furthermore, objections have been raised in many quarters to collection by Government agencies of disaggregated data which violate protection of fundamental civil rights. A case in point is the disagreement with some current Federal agency objectives of equalizing employment by race, religion and sex. Some persons object to the establishment of goals or quotas of this nature, and to the collection and possible evaluation of these data.

Hard-to-measure, subjective data are frequently collected by means of survey research. This technique, according to Albert D. Biderman, has serious limitations which should be remedied before it can be employed on a large scale in social indicators research:

^{131/} Sheldon, Eleanor Bernert and K.C. Land. Social Reporting for the 1970's: A Review and Programmatic Statement. Policy Sciences, v.3, 1972: 141.

...The only currently available means of securing data for indicator systems is through interview and questionnaire surveys. Although the Nation has vast resources for carrying out such surveys, only a small proportion of these conform to the exacting standards of, for example, the Census Bureau. ...Prevailing standards of survey practice are incompatible with meeting even loose criteria of validity... I cannot see the goals of current grant proposals for social indicator systems being met without considerable qualitative and quantitative upgrading of the survey industry ...^{132/}

Notable objections have also been raised to the implementation or policy invocation phase of social indicators activity. Peter Henroit, a widely published student of social indicators, notes that more and better social information does not necessarily lead to better decision-making. Political decision-making, based upon bargaining and compromise of competing interests, occurs, he says, regardless of the level of sophistication of data generated by different decision-making bodies. As a case in point, he reports that:

Public housing was placed where it was in Chicago primarily because of political pressures against dispersion and no amount of information about social costs was likely to have made any difference. ^{133/}

As a variant of this caveat, Frank J. Popper has warned that social indicators, especially the choice of indicators and values upon which they were based, would lead to elitist decision-making far different in tenor from the democratic pluralism characteristic of the United States. As an illustration of this notion, Popper says:

Toward Balanced Growth, the first social report of the Nixon Administration, avoided discussing such issues as national priorities, race relations, urban decay and minority dissent, and its [discussion on] the consumer movement did not mention the activities of Ralph Nader. ^{134/}

^{132/} Biderman, Albert D. Social Indicators -- Whence and Whither. A paper prepared for presentation to the First Annual Social Indicators Conference, American Marketing Association, February 17-18, 1972. [Washington, D.C., unpublished]: p.17.

^{133/} Henroit, Peter J. Political Questions About Social Indicators. The Western Political Quarterly, v. xxiii, no.2, June 1970: 244.

^{134/} Popper, Frank J. The Social Meaning of Social Accounting. Polity, v.4, Autumn 1971: 86.

PART VI. THE POSSIBLE ISSUES OF LEGISLATIVE CONCERN
RAISED BY INFORMATION PRESENTED IN THE REPORT

This report was designed to provide legislative staff with background information on major issues of current concern in program evaluation. Summarized below are the major topics addressed and the possible issues of legislative concern generated by the information presented in the report.

I. The Initiation and Use of Program Evaluation

A. The Need to Clarify Legislative Statements of Program Goals and the Purposes of Evaluation.

1. Consideration of the need for the Congress and for Federal agencies to examine the merits of specifying program goals and objectives more precisely to promote the development of appropriate yardsticks or goals of programs to facilitate the conduct of program evaluation.

2. Consideration by the Congress and especially by authorizing and appropriations committees of the need to clarify and further detail statutory requirements for program evaluation.

3. Congressional consideration of developing appropriate criteria and expectations for the performance of program evaluation, to provide for better correspondence between existing program evaluation techniques and the vague aims of broad-scale social programs, which frequently preclude adequate evaluation.

B. Better Communication for Effective Evaluation.

1. Possible inquiry by the Congress into Federal agency and social scientists' techniques and obligations to provide for improved communications between evaluators and federal administrators to eliminate problems encountered in developing and conducting program evaluations.

2. Possible congressional inquiry into the incentive and reward system of universities which tends to discourage the conduct of program evaluations useful to policy-makers.

C. The Integration of Program Evaluation Into Program Management.

1. Possible congressional inquiry into the extent to which agencies use program evaluation in planning and budgeting and of ways to further enhance the utility of evaluation reports in these activities.

2. Consideration of the need for agencies to develop comprehensive evaluation work plans, setting priorities for program evaluation and recommending program evaluation methodologies most appropriate to the agency's objectives; consideration of a "preassessment of the evaluability of programs."

3. Possible congressional encouragement of the development by agency, of program evaluation handbooks for use within the agency and by State and local administrators who are given greater evaluation and accountability responsibilities with continuing decentralization of some grant programs and revenue sharing.

4. Possible Congressional inquiry into the need for agencies to establish inventories of program evaluation research in process and finished, to assist agencies in eliminating duplicate program evaluation researches, to determine priorities for research, and to assist in promoting development of an overall agency evaluation work plan.

5. Consideration of requiring Government-wide inventories of completed program evaluations to permit sharing of program evaluation information gathered with Federal funds, to enhance the evaluation of similar programs in other agencies and geographic locales, and to stimulate refinement of the methodology of program evaluation.

6. Consideration of the need for social scientists to be encouraged to prepare program evaluation reports in forms most compatible for policy-making purposes.

D. The Need for Better Dissemination of Program Evaluation Information Supported Under Federal Agency Auspices.

1. Consideration of inquiring into the need for Federal agencies to encourage program evaluation information dissemination systems within agencies to assist researchers in sharing information on already completed evaluations.

2. Consideration of Congressional inquiry into the need to establish a Government-wide inventory of abstracted program evaluation reports under auspices of an executive or legislative branch agency or incorporated into existing bibliographic devices to assist researchers in sharing information and to expedite the work of administrators who rely upon evaluations in assessing the merits of a Federal program.

II. Improved Coordination of Program Evaluation in the Executive and Legislative Branches

A. The Need to Incorporate Program Evaluation in Executive Branch Budget Processes.

1. Possible Congressional inquiry into steps the Office of Management and Budget should take to further encourage the use of program evaluation information in preparing budget and program statements, and particularly to encourage Federal agencies to collect hard-to-measure, but relevant, social impact data in the normal course of administering a program to facilitate evaluation of the program.

2. Review, as needed, by the Congress, of steps the OMB is taking to utilize program evaluation information in providing for a better evaluation of how alternative programs help meet national needs, and related to this, the evaluation of the impacts on these issues of the OMB re-organization and "management by objectives" effort.

3. Congressional assessment, as needed, of program evaluation under the revenue sharing program especially with respect to the effectiveness of decentralized accountability, and the quality of auditing and evaluation standards and reporting requirements promulgated by the General Accounting Office and the Office of Revenue Sharing to generate appropriate evaluation information.

4. Consideration by the Congress of proposals to give OMB increased responsibility for centralizing and overseeing executive branch program evaluation activities.

5. Possible Congressional examination of the merits of the recently proposed Federal Program Evaluation Office in the Executive Office of the President or in the Office of Management and Budget.

B. Issues Relating to the Improvement of Legislative Branch Program Evaluation Capabilities.

1. Continued congressional consideration and oversight of ways to fashion a better link between the process of congressional oversight and appropriations to mesh legislative review and evaluation of programs with determination of appropriate program priorities and funding levels.

2. Congressional consideration of establishing a catalogue which would abstract on-going and completed program evaluation studies by agency and social program area, with information in a form compatible with legislative information requirements, either by the General Accounting Office, by the General Accounting Office in cooperation with the Office of Management and Budget, or by a separate unit.

C. Issues Surrounding the Role of the General Accounting Office in Providing the Congress with Improved Program Evaluation Capabilities.

1. Continued liaison between the Office of Program Review and Evaluation in the General Accounting Office and legislative committees to develop statements of program objectives and legislative requirements for evaluation which are compatible with prevailing techniques for the conduct of program evaluation and the collection of information required for evaluations.

III. The Procurement and Conduct of Program Evaluation

1. Possible development of better Government-wide reporting systems, by the National Science Foundation, GAO or OMB, on the conduct and utility of program evaluation, requiring submission of material according to uniform categories treating funding, performer, type of evaluation, purpose, and methodology.

2. Congressional inquiry into the performance of program evaluation research, assessing the relative merits for different types of evaluation of the following institutional performers:

-federally funded research and development centers,

-universities

-in-house, and

-profit vs. non-profit making contractors.

3. Congressional inquiry into problems dealing with funding mechanisms for program evaluation research; and possible revision of statutes, as suggested

by the Commission on Government Procurement, with respect to:

- cost overruns,
- independent research and development,
- sole source contract vs. competitive bidding,
- monopoly of program evaluation contracting by a few industrial firms,
- the merits of using profit-making vs. non-profit-making social science firms or institutions in performing program evaluation research,
- content and impact of poorly stated RFPs for program evaluation research, and
- proper allocation of evaluation functions between in-house and extra-mural performers.

4. Congressional inquiry into the efficacy of designating appropriate levels of funding for program evaluation in authorizing legislation.

A. Alternative Institutions for the Conduct of Program Evaluation Research.

1. Possible inquiry into the merits of establishing program evaluation research units or institutes either for separate agencies, or for a class of social programs to support the conduct and utilization of better social program evaluation research.

2. Specifically, Congressional and Federal agency inquiry into the development of alternative program evaluation institutions which would meet the following concerns:

- independence of administration,
- bringing a national perspective to bear in development of goals and research strategies,
- integration of policies for program evaluation research into other programs and policies for scientific research and development,
- linking of research institutions to client sponsors, distributed throughout the country, and
- provision of adequate training and facilities for interdisciplinary, policy-oriented research.

3. Consideration by the Congress and funding agencies of developing policies and programs to promote the training and utilization of policy-oriented social science researchers, in addition to discipline oriented basic and applied social science researchers.

IV. Methodological Issues in the Conduct of Program Evaluation

1. Oversight, by appropriate committees, of Federal priorities to support

the refinement of program evaluation methodology for use by Federal agencies especially programs supported by the Research Applied to National Science Foundation.

2. Consideration, by appropriate legislative committees and Federal agencies of the pros and cons of social experimentation and of requirements for program evaluations which use a social experimentation methodology. Especially relevant here are the issues of invasion of privacy, ethical and moral questions related to social experimentation, the need to fashion legislative requirements compatible with the conduct of social experiments (including simulation of "real world" conditions, and collection of appropriate continuous data).
3. Consideration of Congressional attention to building into requirements for evaluation of data collection procedures appropriate for program evaluation techniques.
4. Consideration of Congressional and executive branch activities to cross-fertilize program evaluation and social indicators research to strengthen the development of both areas and to assist in providing program evaluators with appropriate and continuous valid social data.

END