

870
12/26/79

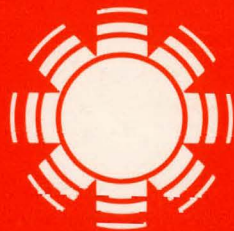
DR. 447

SERI/TR-51-294

September 1979

Solar Energy Program Evaluation: An Introduction

Peter deLeon



SERI

Solar Energy Research Institute

A Division of Midwest Research Institute

1536 Cole Boulevard
Golden, Colorado 80401

Operated for the
U.S. Department of Energy
under Contract No. EG-77-C-01-4042



DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Printed in the United States of America
Available from:
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161
Price:

Microfiche \$3.00
Printed Copy \$ 6.00

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

SERI/TR-51-294
UC CATEGORY: UC-59,60,61,62,63,64

SOLAR ENERGY PROGRAM EVALUATION:
AN INTRODUCTION

PETER DELEON

DISCLAIMER

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

SEPTEMBER 1979

PREPARED UNDER TASK NO. 5328.40

Solar Energy Research Institute

1536 Cole Boulevard
Golden, Colorado 80401

A Division of Midwest Research Institute

Prepared for the
U.S. Department of Energy
Contract No. EG-77-C-01-4042

rb
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

FOREWORD

This report is based on a SERI program task on Program Evaluation Methodology that commenced in July 1978. The purposes of the task were to provide an overview of the practice and methodology of program evaluation and to define more precisely the evaluation techniques and methodologies that would be most appropriate to government organizations which are actively involved in the research, development, and commercialization of solar energy systems. These organizations potentially would include the Department of Energy, other federal departments and agencies, Congress, state and local governments, international agencies, and SERI itself. Explicit in this report are the understandings that quality program evaluation is extremely difficult, that solar energy programs have particular characteristics, and that evaluations need to be tailored with special care to fit the unique characteristics of these programs. These are documented in the body of the report.

This research was initiated by Lewis J. Perelman, who reviewed much of the program evaluation literature and conducted interviews with selected government officials and practiced program evaluators. (A list of the interviewees is appended to the study.) Dr. Perelman prepared and sent out for review a rough draft in late 1978 before leaving SERI to accept a position with the Jet Propulsion Laboratory. At the request of SERI management, Peter deLeon completed the research task by incorporating reviewers' comments and revising the manuscript for publication as a SERI technical report. In responding to the reviewers' critiques, deLeon extensively rewrote the original draft. Perelman reviewed the revised text and requested that he not be listed as an author, citing differences in interpretations and conclusions. Although SERI management regrets his decision, it acknowledges and honors his choice.

Approved for:

SOLAR ENERGY RESEARCH INSTITUTE



James W. Doane, Chief
Policy Analysis Branch

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

ACKNOWLEDGEMENTS

Robert deKieffer offered invaluable assistance and support in every phase of the study. Becky Bryant made important contributions to the early research tasks. I am particularly appreciative of the detailed comments on both the general topic of evaluation and specific suggestions regarding this text offered by Professors Garry D. Brewer of the School of Organization and Management, Yale University and David Kline of the Graduate School of Education, Harvard University. Robert Koontz and Jean Neuendorffer of SERI's professional staff provided incisive technical reviews of this draft. Finally, as noted in the Foreword, this report draws heavily upon the earlier research of Lewis J. Perelman of the Jet Propulsion Laboratory.

Despite the best efforts of these people, I am solely responsible for this report; any error of fact or interpretation must reside with me.

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

SUMMARY

This report is basically directed to two rather distinct audiences whose common denominator is their professional interest in program evaluation. The first audience is composed of program managers for solar energy development and commercialization programs; this audience might be characterized as being technically competent and bureaucratically astute but not necessarily possessing the theory, approaches, and training of the professional evaluator. The second audience, the professional evaluator, possesses this second set of skills while perhaps lacking the technical expertise. The report is written to introduce the problems of each group to the other and to demonstrate how both can gain from a close and complementary cooperation. More specifically, this report argues that formal program evaluation of the individual solar energy programs has great potential for benefiting the national solar energy program and its individual technologies.

Formal evaluation cannot be treated as a single methodological approach for assessing a program. Numerous typologies and approaches for formal evaluation have been advanced, which reflect the evolving nature of the evaluation art and the general lack of consensus among professional evaluators as to theory, methodology, or implementation. In addition to formal program evaluation, informal evaluation—such as conducted by personal observation, Congressional staff, or journalists—is also present. Although it fails to meet the rigorous methodological standards of formal evaluation, it does serve a valuable function; the two should be viewed as complements rather than substitutes.

Key considerations in program evaluation are whether it meets the requirements of the client (in this case, most often a program manager) and if it can be done; i.e., are the necessary data available; will the evaluation effort be supported by management? Evaluability assessment has been proposed as a first step in formal program evaluation; its purpose is to determine if a formal program evaluation should and can be performed. Basically, evaluability assessment attempts to identify the major obstacles to program evaluation before they occur. For example, are the program's objectives sufficiently clear to be translated into program evaluation criteria? If not, then alternative measures must be derived by the evaluation staff and the program manager or else the proposed evaluation might be abandoned.

There are four basic types of evaluation designs, each with its own particular strengths and weaknesses: the pre-experimental design; the quasi-experimental design based on time series; the quasi-experimental design based on comparison groups; and the true experimental design. In ascending order, these basic evaluation design types offer increasing methodological rigor, but this strength greatly increases their cost and reduces their applicability to operating programs. The latter approaches emphasize quantitative methodologies. Recently there has been increased attention paid to the qualitative elements of formal program evaluation, or what some have termed "naturalistic inquiry," in recognition of a number of important qualitative program variables that are not suitable for quantification.

Data considerations are fundamental to any evaluation effort. The initial determination of what data are to be analyzed should be based on a body of accepted theory, but often the evaluator is forced to rely on existing data that may be irrelevant. Additional data questions include the manner in which the data are collected and the effect of those methods upon the data's validity.

The organization and management of program evaluation raises several operational questions. The disciplinary background of the professional evaluator is less important

than the evaluator's ability to delineate the causal relationships that are the basis of program evaluation. The evaluator has to be able to work with people because evaluation is typically a group effort. Evaluators need to balance professional skepticism with the ability to work cooperatively with the program staff, which could present a number of psychological and organizational obstacles to the evaluation of their programs. Evaluations might be requested for the "wrong" reasons, such as to cover up program failures or to "submarine" politically unpopular programs.

The program evaluator must emphasize the value of program evaluation or risk having the study left unread or unheeded. The benefits of evaluation can result in more effective program administration and the improved design of new programs. It must be recognized, however, that there are certain costs to evaluation which need be considered; these costs can be in terms of scarce resources (e.g., financial or staff), the possible reduction of a program's scope, and perhaps even the morale of the program staff. Evaluation should not be viewed as a panacea. It cannot offer objective "truth" nor can it accurately assess the comparative worth of disparate programs, such as health insurance versus housing subsidies. Questions of social worth still are best left to political decision makers.

Solar energy technology programs have grown greatly in their importance and attention in the past five years. These programs differ in a number of critical ways: their position on the basic research to commercialization continuum; their end-uses; and the materials which they utilize (e.g., biomass or wind power). Within the federal government, the lead executive agency for solar energy technologies is the U.S. Department of Energy (DOE); within DOE, the Energy Technology and the Conservation and Solar Application Divisions have the principal responsibility for the development and commercialization of solar energy, respectively. The U.S. Departments of Defense, State, Commerce, Agriculture, and Housing and Urban Development also have responsibilities regarding solar energy technologies. There are a number of state solar energy programs as well; these are mostly regulatory, information dissemination, or subsidization programs and vary widely from state to state. Finally, many solar energy programs, such as land-use requirements, were designed and implemented by local government bodies.

Evaluation of federal programs is, at best, a sporadic and difficult activity. As one study asserted, "There is nothing akin to a comprehensive federal evaluation system. Even within the agencies, orderly and integrated evaluation operations have not been established" (Wholey et al. 1976:15). The main reasons for this lack are an absence of relevant evaluation models, the perceived failure of present bureaucratic incentive structures to encourage program evaluation, the problem of reconciling differing evaluation and information perspectives, and the difficulty provided by minimizing the threats posed to the program by evaluation while maximizing the amount of useful information in evaluation. These obstacles, juxtaposed with the newness and great diversity of most solar energy programs, make the evaluation of these programs difficult even under the best of circumstances. In light of the politicized nature of the solar energy programs and the questionable relevance of earlier technology research, development, demonstration (RD&D), and dissemination models, the evaluation problems become formidable. Some of the earlier evaluation work done on civilian technologies and in the agricultural extension service might be more relevant to solar energy RD&D than the evaluation of military R&D programs or the nuclear power reactor RD&D programs, but their applicability should be carefully examined before these models are used.

Many informal evaluations of the national solar energy programs have been conducted by such groups as the Solar Lobby. Within the federal government, the General Accounting

Office and DOE's Policy and Evaluation Division have evaluated some very specific programs, but mostly on an ad hoc basis. For understandable reasons (e.g., the newness of both solar energy programs and DOE and the inherent problems in evaluating RD&D programs), no coherent strategy for formally evaluating solar energy programs exists; program managers have chosen to place their resources on program design and implementation rather than on retrospective evaluation. There is even less in the way of formal program evaluation on the state level.

The inchoate state of solar energy programs and their objectives combined with the ambiguity of evaluation theory and methodology lead one to conclude that there is no single optimal evaluation strategy, nor should there be. There are, however, important benefits to be gained from program evaluation. For this reason, solar energy program managers should be persuaded of the positive benefits that they can gain from program evaluation. This argument should be made simultaneously by the evaluation community and the highest levels of management within DOE.

In terms of specific policy recommendations, the Solar Energy Research Institute (SERI) can play an important role in solar energy evaluation. Although SERI's mandate does not specify an evaluation role, its unparalleled combination of technical and social science expertise in solar energy make it an ideal candidate for developing solar energy evaluation programs. Furthermore, its close association with DOE program managers and its realization of their particular dilemmas make SERI a uniquely knowledgeable and sensitive reviewer. The key to the evaluation of solar energy programs lies in the extent of cooperation between professional evaluators and the program managers; policies must be formulated and implemented that encourage this relationship. DOE must resolve its current ambiguity regarding evaluation studies, such as which division has the primary evaluation responsibilities. Evaluations of state and local programs must be conducted and the results widely disseminated so that others can benefit from the experiences of these programs; an important feature of the state and local evaluations is that they must be funded at least partially by the pertinent local unit of government so that it retains some commitment to the study and its results.

Underlying these recommendations is the explicit assumption that the formal evaluation of solar energy programs is a worthwhile and valuable endeavor. Although evaluation holds no guarantee for the strict adherence of public policy to its stated objectives or the rational workings of government, it does offer a means by which programs can be measured and a learning experience obtained. Both are imperative if the solar energy technologies are to achieve their promise.

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

TABLE OF CONTENTS

	<u>Page</u>
1.0 Introduction	1
1.1 Audience.....	1
1.2 Organization	2
2.0 Program Evaluation: Roles and Definitions	3
2.1 Program Evaluation in Context: The Policy Process	3
2.2 Definition of Program Evaluation.....	4
2.2.1 Formal Evaluation.....	4
2.2.2 Informal Evaluation	6
2.3 Types of Formal Program Evaluation.....	7
2.3.1 Typology 1: Simple Output	7
2.3.2 Typology 2: Formative/Summative	7
2.3.3 Typology 3: Input/Process/Output	7
2.3.4 Typology 4: Process Levels.....	8
2.3.5 Government Typologies	8
2.3.6 Review	8
2.4 Evaluation Purposes: Three Perspectives	9
2.5 Application of Program Evaluation to Decision Making	9
3.0 Primer on Evaluation Methodology.....	11
3.1 Evaluability Assessment.....	11
3.1.1 Procedures for Evaluability Assessment	11
3.1.2 Development of a Program Model	12
3.2 Basic Evaluation Design.....	13
3.2.1 Considerations Common to all Evaluation Designs	13
3.2.2 Four Basic Evaluation Designs	14
3.2.3 Naturalistic Inquiry.....	16
3.2.4 Comparison of Evaluation Designs	17
3.3 Measurement	18
3.4 Data Considerations	19
3.5 Evaluation Design and Work Plan	20
3.6 Communication of Evaluation Results	21
3.7 Meta-Evaluation: Evaluating Evaluation	21
4.0 Organization and Management of Program Evaluation	23
4.1 Evaluation Personnel	23
4.2 Sponsorship.....	24
4.3 Use of Internal vs. External Evaluators	25
4.4 Timing of Evaluation	26
4.5 Confidentiality	26
4.6 Practical Evaluation Problems	27
4.6.1 Needs and Purposes.....	27
4.7 Program Evaluation in the Political Process.....	28
4.7.1 Reasons Why Evaluations Are Not Done	28

TABLE OF CONTENTS (continued)

	<u>Page</u>
4.7.2	The Psychological Obstacles to Evaluation..... 29
4.7.3	Evaluation for the "Wrong" Reasons..... 29
4.7.4	Ambiguity of Political Decision Making 30
4.8	The Management Environment of Evaluation 30
4.9	The Use of Evaluation Results 31
4.10	The Value of Program Evaluation 33
4.10.1	The Benefits of Evaluation 33
4.10.2	The Costs of Evaluation 33
4.10.3	Limitations of Evaluation 34
4.10.4	The Future of Program Evaluation 34
5.0	Overview of U.S. Solar Energy Programs 37
5.1	What is Solar Energy?..... 37
5.2	Types of Solar Energy Programs 38
5.3	Federal Government Solar Energy Programs 40
5.3.1	Department of Energy 41
5.3.2	Other Executive Departments..... 44
5.3.3	International Programs..... 45
5.4	State Government Solar Energy Programs 45
5.4.1	Financial Incentives 48
5.4.2	Regulatory Measures 48
5.4.3	RD&D and Information Programs 48
5.4.4	State Program Organization 49
5.5	Local Government Solar Energy Programs 49
6.0	The Present Condition of Solar Energy Program Evaluation 51
6.1	Background of Federal Program Evaluation 51
6.2	Characterizing the National Solar Energy Program..... 53
6.3	Federal Solar Program Evaluation 57
6.3.1	Assessment of the Present Condition..... 59
6.4	State Solar Program Evaluation 61
6.4.1	General Description of State Solar Program Evaluation 61
6.4.2	Status of State Evaluation by Program Area 64
7.0	Developing Solar Energy Program Evaluation..... 67
7.1	To Evaluate or not to Evaluate? 67
7.2	The General Context of Solar Program Evaluation 68
7.3	Strategy for Developing Solar Program Evaluation 69
7.4	Developing Evaluation by Program Area..... 70
7.4.1	Evaluation of RD&D Programs 70
7.4.2	Evaluation of Commercialization Programs 73
7.5	SERI Role in Program Evaluation..... 73
7.6	What Should SERI's Role Be?..... 75
7.6.1	Current SERI Activity in Program Evaluation 75
7.6.2	Why SERI Should Do Program Evaluation 75

TABLE OF CONTENTS (concluded)

	<u>Page</u>
7.6.3 Why SERI Should Not Do Program Evaluation	76
7.6.4 Options for SERI Evaluation Role.....	77
8.0 Observations and Recommendations	79
9.0 References	83
Appendix: Consultants.....	89

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

LIST OF FIGURES

	<u>Page</u>
5-1 Types of Solar Energy Programs	39
5-2 Solar Budget: Total Administration Requests and Appropriations, 1971-80	42
5-3 Department of Energy	43
7-1 Solar Energy Evaluation Needs	71

LIST OF TABLES

	<u>Page</u>
5-1 Solar Record, State by State	46

SECTION 1.0

INTRODUCTION

1.1 AUDIENCE

There are two different audiences for this report, audiences whose backgrounds, occupations, and responsibilities are generally distinct from each other yet whose professional intentions can be viewed as quite compatible. Members of the first audience have the responsibility to design and manage solar energy technology development and commercialization programs. These public officials' duties range from basic research on the different solar technologies to the final commercialization of solar technologies. Their responsibilities are more than strictly technical because solar technologies have a number of social, economic, and political ramifications which must be taken into account. The second audience largely consists of professional evaluators; this community has representatives from within and without the government. Still, both audiences share a common interest: that the formal and rigorous evaluation of government programs, in this case solar energy technology programs, can make these programs more efficient and effective in achieving their objectives than might be the case presently.

At the risk of oversimplification, the first audience can be characterized as being technically competent and politically sensitive but not necessarily possessing the theory, approaches, and techniques of the skilled program evaluator. Similarly, the professional evaluator probably lacks the technical competence and bureaucratic sensitivities of the program manager who is charged with developing a given solar energy technology. These characterizations are not totally accurate, of course, nor are they meant in a malicious manner. Rather, the characterizations are meant to suggest that the two audiences for this report do not share an identical set of skills, needs, or perceptions. The explicit purpose of this report is to bridge whatever gaps do exist so that the two may appreciate each other's particular problems and cooperate toward the careful, quality evaluation of solar energy programs with a minimum of intellectual dissonance.

We have been talking in general terms of two audiences. It is useful at this point to disaggregate these audiences and discuss more precisely who, in fact, the prospective audiences for this report are. The first, which we have broadly described as solar energy program managers, is composed of those persons who have strategic and operational responsibilities for developing, demonstrating, or disseminating various types of solar energy technologies. These people are situated on the federal, state, or local government levels. Obvious candidates include members of the Department of Energy, members of Congress and their staffs, state energy officials, and local building code personnel. This hierarchy is detailed in Section 5.0, Overview of U.S. Solar Energy Programs. The evaluation community is equally dispersed and perhaps even harder to typify. It includes representatives from universities, consultant groups, private industry, research organizations, and professional evaluators from the different government bodies (such as the staff of the U.S. General Accounting Office or the Policy and Evaluation Division of the Department of Energy).

In attempting to bridge the intellectual and professional gaps that might exist between these two communities, this report inevitably must satisfy neither of the groups. The description of the national solar energy programs will appear to the solar energy program manager as somewhat simplistic, not fully reflecting the true complexity of the task that makes his or her position and perspective resistant to objective analysis. At the same

time, the overview of evaluation techniques and the discussion of what is needed to produce a quality program evaluation will seem trite to the professional evaluator. However, each audience needs to remember that the purpose of this report is to provide the basic information groundwork for the other; i.e., to inform the solar program manager of the costs and benefits of evaluation and the professional evaluator of particular evaluation problems unique to solar energy programs. By serving these two education functions, this report will facilitate communications between the solar program manager and the professional evaluator. This will result in early and accurate expectations from both parties as to how they might cooperate in providing solar energy programs that are more consonant with the program objectives and in effectively meeting the overarching national priorities. Although neither audience might be completely satisfied with the level of detail found in this report regarding its particular area of expertise, each audience will gain enough from other sections of the report such that the audiences' exchanges and appreciations of each others' problems and requirements will compensate for any disappointment.

One of the principal reasons for preparing this report is to demonstrate that formal program evaluation of solar energy programs has great potential for benefiting the national solar energy program and its individual technologies. At the present time, this rationale must stand as little more than assertion, although we plan to argue persuasively about its validity, for there are few, if any, examples of formal, rigorous evaluations of solar energy programs. This lack is explained by perfectly justifiable reasons discussed in the text of the report. Indeed, there are precious few examples of evaluations of broad-gauge energy policies, if one excludes the very political charges that policy X has "not worked."* For this reason, many of the illustrations used in the text come from subject areas outside of solar energy, such as mental health programs, social welfare programs, military R&D, and, especially, education. Clearly, education and social welfare are not solar energy; therefore, the necessary caveats must be recognized and observed. Still, even granting the special features of solar energy technology development, important parallels can be suggested concerning program evaluation from areas that are not energy related.

1.2 ORGANIZATION

The report is organized so that the reader is first introduced to the role and issues of evaluation. This is to provide a set of issues to organize the subsequent sections detailing the national solar energy programs. Then, these two themes are integrated by examining the evaluation strategies and methodologies tailored to fit the particular needs of the various individual solar energy programs. Specifically, this report is divided into eight sections, including this introduction. Section 2.0 delineates the place of program evaluation in the overall policy process and offers a number of definitions used throughout the report. Section 3.0 is a basic primer in evaluation methods and difficulties, while Section 4.0 discusses the resources necessary for evaluation as well as their limitations. Section 5.0 briefly reviews the national solar energy public programs. Section 6.0 examines the current efforts to evaluate these programs. Section 7.0 assesses SERI's potential role in solar evaluation, and Section 8.0 offers specific recommendations as to what might be done to apply evaluation techniques to solar energy programs.

*An important exception to the proported lack of general energy evaluations is the literature that reviews quantitative energy models. See, for example, Limaye, ed. (1974); Brock and Nesbitt (1977); and Searl, ed. (1973).

SECTION 2.0

PROGRAM EVALUATION: ROLES AND DEFINITIONS

The general research activity of measuring and judging the performance and effects of public programs is referred to by a great variety of names, such as program evaluation, evaluation research, auditing, implementation review, monitoring, program assessment, and policy analysis. This semantic proliferation is more than merely inconvenient; it can result in a failure to understand, properly use, or assign appropriate responsibility for what could be a critical research function with an important policy payoff. This report will use the terms "program evaluation" or "evaluation" to refer to the specific activities defined below.

2.1 PROGRAM EVALUATION IN CONTEXT: THE POLICY PROCESS

Before turning our primary attention to program evaluation, it would be useful to present evaluation in the general context of the policy process. That is, where does program evaluation fit into the making and execution of public policy? In the absence of such a context, evaluation might easily be viewed as an isolated, disjointed activity rather than an integral and integrated part of the policy process. This is not to say that evaluation cannot be extracted and scrutinized; rather, it is initially important to understand its role in the overall policy process before turning to the specific study of evaluation.

Drawing upon the seminal work of Lasswell (1971: Chapter 2), Brewer has proposed a six-step model of the policy process (1974):

- Initiation/Invention
- Estimation
- Selection
- Implementation
- Evaluation
- Termination

Very briefly, the first step consists of the individual and organizational recognition of a societal problem, the definition of the policy objectives in response to the problem, and the generation of possible policy options. The estimation stage weighs each alternative's costs and benefits against the general policy objectives prior to the selection stage, in which the decision maker chooses among the policy options. The policy is translated into specific programs and carried out during the implementation stage. In the evaluation stage, the effects of the program relative to the policy objectives are assessed as well as any unanticipated consequences that the policy and its constituent programs may have produced. Finally, the termination state is a recognition that a specific policy needs (or should) not to be continued indefinitely; when it can be demonstrated that a policy's objectives are reached and maintained, its relevance and applicability should be reconsidered and terminated if found redundant, outmoded, or dysfunctional. (This depiction of the policy process is elaborated upon by Brewer and deLeon, forthcoming).

It should be stressed that this is a highly iterative model and that none of the stages described should be treated as self-contained or exclusive of the others. The initiation

stage easily could be the result of an evaluation of a program that was not meeting its prescribed objectives. Estimation of the policy options fundamentally affects the selection process, just as implementation analysis does (Hargrove 1975; Bardach 1977). And, program termination is as much a policy option as the initiation of a new program (deLeon 1978).

Within each of these stages, a number of disciplinary and interdisciplinary tools or approaches can be utilized. These fall under the general rubric of "policy analysis" and include such disparate approaches as statistical analysis, organizational theory, economics, survey research, operations research, demographics, decision analysis, computer simulation, political science, and law. All of these approaches or analytic tools can be applicable, but the key to their specific utility in a given situation is the requirements of the particular task at hand. Linear regression techniques might be completely appropriate for evaluating the effectiveness of new teaching methods and totally inappropriate for designing implementation strategies. Simply, there is no single best approach to policy analysis.

2.2 DEFINITION OF PROGRAM EVALUATION

A policy may be any purposeful set of activities undertaken by a government with the objective of satisfying certain broad societal objectives. Programs are the various instrumentalities created to realize policy objectives, and projects are the fine-grained, tactical devices undertaken within the general programmatic framework.* For example, an energy-related policy would be to reduce U.S. dependence on foreign petroleum (see deLeon and McNown 1979); an attendant program would be the reduction in the American consumption of gasoline so a project might be the rationing of gasoline by closing service stations on Sundays or, in the case of solar energy, promoting gasohol by waiving the federal tax on it.

2.2.1 Formal Evaluation

Program evaluation is a formal process of studying programs that contains these necessary characteristics:

- an explicit description of what the program actually is doing, has done, or was intended to do;
- some systematic measurement of the effects of the program is used; the criteria and data employed are agreed to and dependable;
- a comparison of the measured effects of the program to the defined or ascertained set of program objectives is present; and
- the potential in the application of the resulting information to policy or program decisions is real.

A number of points need to be added to clarify this definition of program evaluation. First, the basic perspective of program evaluation focuses on what is happening or has happened. Although program evaluation can, and indeed should, have an effect on the

*Wholey et al. (1976) draw a similar set of distinctions.

design and implementation of future policies and programs, studies of future options, policies, costs, impacts, or outcomes are not included in the present definition of program evaluation (Hatry et al. 1973: 8).

Second, the measurement of program effects may be qualitative to a significant extent, using scales, indices, or ordinal measures based at least partly on subjective judgment. However, like its quantitative kin, the qualitative measures should be systematic, reasonably understandable, and at least nominally reproducible. Also, program outcomes and effects should be distinguished and evaluated separately. Outcomes are the immediate results of a program in terms of consequences that can be measured and associated with purposive acts. Effects can and should go beyond the intentional outcomes of the program to include evidence of program processes and of both positive and negative unintended consequences, or "externalities" (Rutman, ed. 1977: 17). The latter are generally more long-term in their nature.

Third, evaluation does not assign values to a given program but rather tests a program against an explicit set of given values or criteria (U.S. Comptroller General 1978: 1-5). The explicit purpose of program evaluation is not to label a program as "good" or "bad," but it is to demonstrate how the results of the program compare with specific objectives, expectations, standards, or conditions.

Fourth, the criteria against which a program is evaluated are not confined to formal statements of program goals and objectives, but they also may include the values, goals, interests, and needs of any stakeholder. Indeed, part of the evaluation act is to specify precisely what the programmatic objectives are. Formal statements of goals commonly are highly ambiguous or may be absent altogether; in many cases, the relationship between the nominal goals and the actual purposes that the program is intended or expected to serve is obscure or even dubious.

Fifth, a basic assumption of this study is that there is no single, ubiquitous evaluation model; that different forms, concepts, and approaches are needed to satisfy different programmatic requirements and objectives. One useful typology distinguishes among Process, Response, and Impact Evaluations (see Brewer and deLeon, forthcoming, Chap. 6). Process evaluation concentrates on the internal workings of an institution. Probably the most common form of evaluation, it includes management audits, personnel reviews, informal assessments of performance, and other routine appraisals of "how are we doing?" As such, it essentially is found in any organization. Response evaluations concern programs that an institution designs or manages in response to external stimuli or demands. Finally, Impact evaluation focuses on the differences that an organization and its programs make on its surrounding environment. The Head Start programs provide a set of examples. Head Start was prompted by legitimate worries that minority children often were being disadvantaged by the educational system because they were not able to start on an intellectual or emotional footing equal to their white classmates; Head Start programs were designed as a response to these perceptions. A response evaluation of Head Start, therefore, would be directed to assessing how the Department of Health, Education, and Welfare (HEW) responds to the demands by minority groups that these learning impediments should be minimized. An impact evaluation of Head Start, on the other hand, would have evaluated how effective the various Head Start programs were in eliminating the performance differential between the two groups of students.

This report primarily addresses the response and impact types of evaluation because, as Hatry et al. note, "While a program evaluation may include consideration of workload measures, operating procedures, or staffing, its chief focus is on measuring the program's impacts or effects. Evaluation aimed solely at a program's internal procedures, staffing,

and management might better be labeled management evaluations or organizational audits" (1973: 8). The types of evaluation commonly required by management-by-objectives, zero-based budgeting, management audit, or similar procedures are not treated as program evaluations in this report.

Finally, it should be recognized that the present state of evaluation theory is not well developed and that little consensus exists; the terminological, classificatory, and methodological babble that characterizes the literature merely reflects this fact. It is critical, therefore, that the program evaluator realize this condition and make his or her own translations and equivalencies of the necessary terms and requirements when any given evaluation study is confronted.

2.2.2 Informal Evaluation

The above discussion looks at program evaluation as a formal research process. In reality, however, programs commonly are "evaluated" in informal ways that satisfy few, if any, of the above criteria. Investigative journalists, public interest groups, lobbyists, commercial associations, private foundations, universities, legislative committees, public hearings, and lawsuits are some of the mechanisms by which government programs often are "evaluated" (Hatry et al. 1973: 13-14). These means of critical program assessment frequently make up for a lack in scientific validity by having a real impact on government decision making. Although such activities have their place in the political or management processes, they provide little reliable, consistent information by themselves about the effects of public programs. Despite the occasional revelation of corruption, malfeasance, or general program deficiencies, these informal evaluation approaches have several inherent shortcomings. They tend to focus on the inputs or process of managing the program rather than the results. They generally address the spectacular or politically sensitive issues rather than the general level and quality of services. The procedures used to gather information are often unsystematic and biased. The shortage of time and manpower provided for program review and the immediacy of the tasks vitiates meaningful, thorough evaluations. Finally, they seldom have the perspective to determine whether the results of the program were beneficial in the long term. Thus, the information presented may not represent a balanced assessment of the true conditions (Adopted from Hatry et al. 1973: 15-16).

More rigorous approaches to management audit (e.g., PPBS, MBO, ZBB) may provide more thorough information about program effects, though not of the depth of formal evaluation processes. Nevertheless, such systems usually require a periodic comparison of measured program outputs with projected goals. Although the procedure need not be very costly, there is growing evidence that such systems are neither cheap nor easy to implement or use (see Novick 1972; Wildavsky 1974: Chap. 6). None of the above activities is an adequate substitute for rigorous evaluation. However, they may provide valuable information that can be useful in the identification of the need for a program evaluation, in the evaluability assessment, and sometimes in the evaluation itself.

Program managers seek and receive feedback from a variety of informal, often personal sources. The limitations of this informal feedback constitute the chief constraint on a program manager's ability to manage effectively and on a policy maker's ability to exercise effective oversight of policy implementation (see Bardach 1974).

If "evaluation" is considered most simply as feedback on the effect of government programs to policy makers, informal evaluation mechanisms are clearly the most common,

most inexpensive, and (probably) most influential forms of evaluation. The major deficiency from the standpoint of the professional evaluation is the high degree of subjectivity; the reliance on information which is inaccurate, unreliable, and sometimes simply dishonest; and the bias of the conclusions. This is not to imply that formal evaluations are immune from such charges, only that they should be less prone.

Informal evaluations serve as important functions. They can alert a decision maker that irregularities that require formal evaluation are occurring; they can verify formal evaluation findings, especially those of qualitative nature; and they can often attract the public attention that eludes formal evaluation and that precedes political action. Program evaluation provides an additional, higher resolution information channel which should produce more thorough, accurate, and relevant evidence about a program's performance than informal evaluation media provide. Neither channel should supplant or replace the other; the two should complement and enrich each other (see Downs 1967; Wilensky 1967). Indeed, they should be structured to enhance their complementarity.

2.3 TYPES OF FORMAL PROGRAM EVALUATION

Formal program evaluation includes a variety of distinguishable research activities. Several classifications exist to identify different types of program evaluation. Following Kline (1978: IX-9 through IX-13), some of the major typologies are:

2.3.1 Typology 1: Simple Output

The simplest typology views evaluation only as output evaluation. This type of evaluation reviews only the ultimate effects of programs without consideration of objectives, needs, processes, or inputs.

2.3.2 Typology 2: Formative/Summative

Michael Scriven (1967) posits a typology that distinguishes between formative and summative evaluation. The purpose of formative evaluation is to provide information to decision makers, which should improve the quality of the program. An underlying assumption of formative evaluation is that program actions are incremental in nature so that modifications in the program can be implemented and can result in the removal of existing programmatic problems or deficiencies. The intent of summative evaluation is to evaluate the overall effectiveness or value of a program. The resulting information is not oriented primarily to the operational program manager but to policy makers concerned with the continuation or termination of the program, the assessment of success or failure for historical or political purposes, or the application of the lessons of one program's experience to other areas of government activity.

2.3.3 Typology 3: Input/Process/Output

Daniel Stufflebeam's (1971) evaluation typology focuses on three kinds of information needed for program evaluation: input, process, and output. Input information describes events or resources that are required for the initiation of program activity; for example, instructional materials and appropriately equipped workshops for a program designed to train solar equipment installers. Process information concerns the events that are taking

place while the program is underway, and describes the procedure by which the program pursues its intended results. Process measures of the solar installer training program might be the number of qualified instructors or class enrollment. Output information describes the effects of the program. Output information that would be of interest in evaluating the training program for solar equipment installers would include, for example, the graduation rate and the proportion of graduates employed as equipment installers.

2.3.4 Typology 4: Process Levels

A fourth evaluation typology, suggested by Edward Suchman (1967), focuses on the evaluation process and includes five types of evaluation—effort, performance, adequacy of performance, efficiency, and process—that roughly reflect the cost and utility of the resulting information.

Effort evaluation measures the quantity and quality of program inputs, while performance evaluation relates measured outputs to specified program objectives. Adequacy of performance evaluation further links measured performance to some indicator of the amount of need for the program output. Efficiency evaluation estimates program outputs in terms of monetary benefits or indicators of effectiveness to total costs. Process evaluation attempts to determine the functional relationships between program causes and effects. In general, the difficulty of these five types of program evaluation increases in the order they are presented; i.e., effort is the easiest and process is the most difficult.

2.3.5 Government Typologies

Various government agencies have established their own typologies of program evaluation as a matter of standard procedure. For example, one of the first federal agencies to undertake formal program evaluation, the Office of Economic Opportunity (OEO), officially established three types of evaluations. The first entailed an overall assessment of program effectiveness where the emphasis was on determining the extent to which the programs were successful in achieving basic objectives. The second evaluated the relative effectiveness of different program strategies and variables where the emphasis was on determining which of the alternative techniques were most productive for carrying out a program. The last emphasized the assessment of managerial and operational efficiency (U.S. OEO, 1968). This typology largely was adopted in DOE's first interim procedures for program evaluation. The Appendix to Interim Management Directive (IMD) 0203 defined four types of evaluations: effectiveness, performance, efficiency, and audit.

A concern reflected in both these examples but not indicated in the academic literature is the need to associate different types of evaluation with the appropriate levels in bureaucratic hierarchies. This reflects the fact that different levels of decision making have different information requirements.

2.3.6 Review

This sampling of typologies is useful in demonstrating the inchoate state of evaluation theory, the variety of reasons for undertaking evaluation, the diversity of possible approaches, and the need to design a particular evaluation program strategy to match a

program's requirements. The problem with such typologies, especially when they are embedded in government regulations, is that they risk becoming procrustean. They can assume a rationale of their own, forcing evaluation tasks to fit into a rigid classification scheme rather than reflecting the flexibility that should be inherent in good evaluation design. The broader the range of program activities to be evaluated, the less likely that any fixed typology can accommodate the variety of evaluation strategies required. This is especially true of new programs and agencies because of their newness. For example, commenting on DOE's IMD 0203, Tom Glennan has noted that

[The] formidable array of [DOE] program types means that DOE evaluation policies must be very flexible in order to be relevant. . . . The meaning of evaluation, the techniques for performing evaluation, and the nature of the performers of evaluations will vary considerably among these program types. . . . [T]he typology of evaluations which is provided in the Appendix to IMD 0203 cannot be applied to all of these program types and, as a consequence, provides poor guidance to those charged with planning an evaluation program. (Glennan 1979)

Recognizing the great diversity of solar energy programs, this report neither recommends nor relies on any one typology of evaluation. Rather, types of evaluation from any of the above classifications are referred to as appropriate in particular situations.

2.4 EVALUATION PURPOSES: THREE PERSPECTIVES

Program evaluation has multiple clients and, hence, many objectives. In her analysis of the Symposium on the Use of Evaluation by Federal Agencies, Chelimsky (1977: 6) identified three general perspectives on the purposes of evaluation. The first is a knowledge perspective that proposes that evaluation should establish evidence leading to new insights about social problems and about the effectiveness of governmental strategies for addressing them. A second view adopts a management perspective to use evaluation as a flexible management tool—a support system for assessing and improving the operational efficiency of government programs. The third is an accountability perspective that treats evaluation as the best use of tax resources by holding program managers accountable for the worth (i.e., both the effectiveness and operational quality) of their programs.

The three perspectives are all valid and should not be seen as mutually exclusive. Still, they are not necessarily complementary. The high reliability and internal validity implied by the knowledge perspective may be too costly or time consuming to meet the near-term demands of the management perspective. The management perspective may imply a more casual approach to evaluation and be at odds with the accountability perspective's demand for public examination of bureaucratic performance. Or, the accountability perspective's expectation of simple, aggregate appraisals of a program's merits usually contradicts the knowledge perspective's desire for precise, even elegant, data interpretation.

2.5 APPLICATION OF PROGRAM EVALUATION TO DECISION MAKING

A few general points may be made concerning the way in which program evaluation can be applied to decision making. First, effective program evaluation nearly always must go beyond the identification of outcomes and effects to describe how these results were obtained. Good evaluation not only analyzes success or failure but also explains why a

particular program has been successful or unsuccessful. This suggests the importance of evaluating processes as well as outcomes. Indeed, many types of public programs (e.g., some educational programs and R&D programs) must be evaluated primarily in terms of processes because ultimate outcomes are diffuse, delayed, or otherwise difficult to attribute to program activities.

Second, decision makers generally want evaluation to assess a range of alternatives that they can affect. Yet, traditional program evaluation usually generates information which is more appropriate to the simple decision of whether to continue or terminate an existing program. Program evaluation is limited in its ability to appraise a wide range of decision options because of the difficulty of extrapolation from one set of experiences to another. To the extent that existing program activities are highly diverse, good evaluation may illuminate the relative value of alternative program strategies. In reality though, few programs are implemented in a form that satisfies the requirements of good experimental, or even quasi-experimental, evaluation designs. As a result, few evaluations can provide conclusive information about the effectiveness of a wide range of policy options, as opposed to persuasive evidence regarding specific programs. Those responsibilities are part of what was described earlier as the estimation process.

SECTION 3.0

PRIMER ON EVALUATION METHODOLOGY

Evaluation subsumes a number of component processes which must be addressed in a particular program evaluation study. This chapter provides a brief introduction to those processes for those who may be unfamiliar with the essential mechanics of program evaluation. It considers such issues as evaluability assessment, evaluation design, data requirements, and the evaluation of evaluation.

3.1 EVALUABILITY ASSESSMENT

The concept of "evaluability assessment" has been proposed by Nay and Kay (1978) as a first step in program evaluation. The overall purpose of evaluability assessment is to decide whether to evaluate, what to evaluate, and how to proceed with the evaluation.

In the past decade, large-scale program evaluations, which often were technically complex, costly, time-consuming, and disruptive, resulted in little positive effect on government decision making. This experience led to a widespread image among program managers of evaluation as expensive, slow, and impractical. Evaluability assessment was developed as a response. The essential approach is to define the information requirements and to structure research to provide incremental information appropriate to the various levels of decision making.

Evaluability assessment serves two purposes. The first is quickly and cheaply identifying the questions that further evaluation research can answer. This implies that "the purchasers of any evaluation would be better advised to buy information in relatively small, sequential steps rather than gamble that one early, predetermined evaluation design will result in a useful set of answers to their questions" (Nay and Kay 1978: 245). Second, evaluability assessment is designed to focus on the most immediate interests and concerns of the program managers. In contrast to the problem of getting managers to consider the results of many evaluation studies, evaluability models are intended to draw the attention of even high-level decision makers. The evaluability approach is designed to illuminate discrepancies between rhetoric and practice, often giving decision makers early insights into what the real problems of the program are.

Evaluability assessment may be useful even when it does not result in an evaluable model and implementation of the evaluation design. A decision to postpone an ill-framed evaluation, modify the design or implementation of the program, or revise the expectations or objectives of program managers are equally valid and important results.

3.1.1 Procedure for Evaluability Assessment

Many considerations enter into the decision of whether a particular program should be evaluated. Too often this decision is based on incomplete understanding about what information a full-scale program evaluation can and should provide. Evaluability assessment tries to answer these questions once a potential program for evaluation has been designated. Three general criteria may be considered: program significance, evaluation effect, and evaluation feasibility (Hatry et al. 1973: 110-111).

Program significance determines whether a program is significant enough to warrant formal evaluation. It considers the requisite investment of resources and the potential need for a higher level of resources. There should be evidence of important costs and benefits attached to the programs as well as an indication that they may be performing marginally and they may be amenable to major improvements.

A second set of criteria weighs the potential effect of the evaluation on decision making. This considers the absence or presence of preconceptions or prejudices by decision makers or of political pressures, commitments, or constraints that would preclude any effective response to evaluation information. It also encompasses major decisions pending in regard to the continuation, termination, or modification of the program that are expected to be made before the evaluation can produce useful results.

The evaluation feasibility criteria essentially anticipate major obstacles to performing an evaluation, such as the availability of adequate data for the evaluation tasks. Similarly, the question of sufficient resources to satisfy the technical requirements of the evaluation is broached. Institutional problems include the political and organizational support to ensure that the evaluation can be conducted without excessive interference and sufficient program stability so that the program will not change substantially before the evaluation is completed.

3.1.2 Development of a Program Model

Evaluation must be based on some model or description of what the program actually has accomplished. Often, evaluation studies are based on inarticulated and undocumented models of the program retained in the minds of the evaluation clients and the evaluators. The problem does not exist necessarily with the inaccuracy of the mental model; much more critical is its implicit nature that often inhibits those concerned with the evaluation from sharing their assumptions and expectations. Many of the historic failures of program evaluation can be attributed to ambiguity in the mental models of evaluators and clients representing the behavior of the program, what evaluation information is needed, and how the evaluation produces the information. Evaluability assessment generates a model of the program which is explicit and accepted by both clients and evaluators.

Evaluability assessment develops a model of program behavior by simultaneously gathering rhetorical descriptions and observing the actual programs. The researcher interviews program participants and others with some knowledge of the program to obtain descriptions of what various parties think the program is doing. The researcher or another member of the research team also observes the behavior of those involved in the program to determine its effect from a relatively impartial viewpoint. The gathering of rhetorical descriptions and the observation of actual behavior theoretically are carried out systematically throughout the program by following the chain of command, the flow of funds or products, or some other unifying factor; although institutional obstacles often intrude (see Elmore 1978). The researcher then synthesizes the information into two classes of models, a testable model and an equivalency model. The testable model of the program represents the researcher's formulation of the rhetorical descriptions; i.e., what people say the program is accomplishing. The model is "testable" in the sense that it constitutes a set of observations about what is happening that can be tested in some verifiable manner. The equivalency model of the program represents the researcher's observations of what actually is taking place based on what the program participants were observed to be doing rather than on what they said they were doing. "Equivalency"

means that the model can be considered equivalent to reality for the purposes of the evaluability assessment.

The next stage in evaluability assessment is the analysis and comparison of the two models. First, the models are examined to determine whether the internal structure is plausible, whether relevant parts are missing, whether the model is self-consistent; in sum, their general validity. Next, comparisons are made between the models and their source material to see whether the information has been represented accurately. These comparisons insure that the testable model does represent the rhetorical descriptions of those interviewed and that the equivalency model is consonant with the observations.

The last stage of the evaluability assessment compares the testable model with the equivalency model to identify the discrepancies between the rhetoric and the reality of program behavior as seen by a relatively neutral observer. Assuming no unanswerable questions or insufferable obstacles are raised, the differences between the testable and equivalency models are reconciled to produce a single "evaluative model" that serves as the basis for the design of a formal evaluation process. The evaluative model represents an agreement between the program managers and the evaluators about what is being evaluated, how the evaluation is to be done, and what is to be accomplished.

3.2 BASIC EVALUATION DESIGN

Evaluation design is a technically demanding task. This section provides a brief synopsis of some of the fundamentals of evaluation design. For greater detail, see Hatry et al. (1973), Wholey et al. (1976), and van Manner (1973).

3.2.1 Considerations Common to All Evaluation Designs

The basic problem of what we have defined as impact evaluation is to identify effects; i.e., events which can be attributed to the action of the program. At the limit, this problem is insoluble. No evaluation measurement can prove beyond question that a government program caused a particular event that would not have occurred if the program had not existed. However, several alternative designs of evaluation research can indicate with varying degrees of confidence that something occurred as a result of a program that would not have occurred otherwise. The best of these designs yields information that merits a high level of confidence. Evaluation cannot eliminate uncertainty in government decision making, but it can be significantly reduced under some circumstances.

A formal evaluation should measure a program's performance against a single criterion in only the most extenuating circumstances. Criteria of program evaluation should reflect the major concerns not only of the evaluation client but of all the important constituents in the program. An effective evaluation not only reflects the client's immediate interests but prepares the client to respond to the demands of other parties interested in or affected by the program. Positive and negative externalities must be considered. Some evaluations go beyond the immediate issues to anticipate important issues about the program that may be emerging.

Regardless of what design or method of evaluation is used in an evaluation study, a thorough search must be made for plausible explanations of measured changes other than the program interventions. Coincidence never can be ruled out completely as the cause of a program's apparent outcome. Numerous alternative explanations of measured effects often will exist that need to be examined and accounted for in the final report.

Evaluation often is supported with the expectation that the results can be generalized, but their generalization can be both important and frustrating. The temptation is great to claim that program X produced certain results in a particular situation, therefore, (1) program X will yield similar results when duplicated in a different situation (e.g., another city); (2) program Y will yield similar results if modified to make it similar to program X; (3) program X will produce similar but greater results when increased in scale to program X+; or (4) program X will continue to produce the same results in the future. Unfortunately, such extrapolations rarely are valid. As Kline (1978: 13) notes, evaluation differs from basic social research mainly in that it is designed to answer specific programmatic questions rather than to support broad generalizations.

Another general problem is the choice of the time horizon over which a program's effects are to be evaluated. There is no inherent reason why the value of a program should be judged by the effects demonstrated over a period of six months, five years, or even ten years. The time span over which the results of a program are measured inevitably is dictated by the resources available to carry out the evaluation, the timeliness of the resulting information, and the interest of the sponsor. Many public programs produce critically important benefits and costs that may not become evident until many years after the program is initiated or even terminated. On the other hand, there are important reasons why a program should operate for a certain period of time before being subjected to a formal evaluation. The New Jersey income maintenance experiment is a program whose premature evaluation was driven by political expediencies and whose evaluations, therefore, were suspect (see Rivlin 1971).

This problem of long-term effects is particularly perplexing in the evaluation of technology R&D programs. Not only are the benefits of R&D usually long-term in nature, but the results of an R&D program can appear negative right up to a breakthrough or turning point if the development of an innovative technology is the program goal. Cost overruns and schedule slippages are short-term standards that may be secondary considerations in comparison to the longer term benefits of the R&D process. The delayed effects problem may be particularly pertinent to solar energy programs. The overall development and diffusion of solar energy systems can have high "front-end" costs and still produce significant but delayed benefits. Relatively short-term evaluations could underestimate the long-term value of solar energy programs. Alternatively, evaluation of conventional energy resource programs, which compete with solar energy, may demonstrate comparatively greater near-term benefits while neglecting possible long-term costs, such as resource depletion or environmental damage.

A final general problem, endemic with most evaluations, is using data not intended for evaluation. Prospective evaluation designs specify procedures for gathering measurement data while the program is in progress. Ideally, the evaluation design is completed and the procedures for data collection are established before the program has been initiated. In practice, however, prospective designs are not always possible and the evaluation must depend partly or entirely on existing evidence about the program's performance. Though such retrospective evaluations are often necessary and may prove valuable, prospective designs are preferable. The key around this dilemma is to pay heed to program evaluation while planning the program, in other words, well before the evaluation is initiated.

3.2.2 Four Basic Evaluation Designs

Though many different evaluation designs are possible and every evaluation study in operation represents a unique design, we can identify four basic types of evaluation

design: pre-experimental designs, the two types of quasi-experimental designs, and the true experimental design (Campbell and Stanley 1963).

The pre-experimental designs are characterized by the observation of program outcomes after program actions have occurred and by the absence of any formal control samples to contrast with the groups to be observed. The most robust of the pre-experimental designs is the one-group, pretest, posttest design: An observation (measurement) of the evaluation criteria among the program target groups is made prior to the program action; the program action is implemented; and another observation of the evaluation criteria is taken to see what change has occurred.

A second pre-experimental design is the static group comparison. Here, observations of two groups are made after the program action has been implemented; one group that received the program treatment and another that did not. The evaluation criteria are compared between the treatment and the nontreatment groups.

From a policy perspective, the weakest form of pre-experimental design, yet the one that may represent the most common form of program evaluation, is the case study in which the evaluation criteria are determined and applied to the target group after the program is already ongoing. Although case studies can be extremely valuable, their individualistic nature vitiates the applicability of their findings to other situations.

The pre-experimental design is the simplest and most available. Limited time and staff may make this design the most practical approach to evaluation. Such an approach may be warranted when the program or project is small scale and of short duration. Data collection and technical skill requirements usually are minimal. The results of this type of evaluation also are the least valid in terms of demonstrating that specific effects resulted from program activities.

Quasi-experimental designs employ more rigorous quantitative means (e.g., statistical inference) to induce the effects attributable to program actions than those just discussed. There are two basic quasi-experimental designs: time series designs and comparison group designs.

The time series design uses a longitudinal series of observations of the evaluation criteria in a single target group made at several fixed time intervals both before and after the implementation of the program. Time series experiments compare trends in the criteria before and after the program action. A variation is the multiple time series design which uses simultaneous time series observations for both treatment and nontreatment groups. The time series design is appropriate when the program is expected to affect behavior (i.e., a definite change in a criterion variable). For instance, data for an energy conservation program might consist of time-series data showing a change in the rate of energy consumption in a target community following the program's introduction. Erratic or cyclical behavior in the criteria variables may necessitate careful statistical analysis; statistical methodologies have been developed specifically for these purposes. Under favorable circumstances, the validity of time-series data may be quite high. The cost is moderate depending on the acquisition of data and the technical skills required for data analysis.

The nonequivalent comparison group design applies the evaluation criteria before and after the program action has been implemented among both treatment and nontreatment groups. Groups are chosen that are as similar as possible (e.g., two cities with similar demographic and socioeconomic characteristics) but are treated as nonequivalent

because selection of the groups is not statistically random. This design compares the effectiveness of several alternative program strategies by identifying a separate target group for each strategy and assessing the evaluation criteria in each group before and after program implementation. The nontreatment group also is measured (see Klitgaard 1978).

The comparison group design eliminates coincidental factors that otherwise might be attributable to program activities. For example, it may be that the region where the energy conservation program was introduced had an unusually mild winter, thus, making it questionable whether reduced energy consumption resulted from the program or simply from lessened demand. If a similar nontreatment community existed in the same meteorological region as the treatment community, data from both communities may indicate whether the conservation program affected energy consumption differently than did the mild weather. Identification of a comparable community, though, may be difficult. Even if a comparable community exists, comparable evaluation data may not be available. Costs of this design are moderate.

The true experimental design uses randomly selected test and control groups to insure that the groups are as statistically equivalent as possible. The basic design is the pretest/posttest, control group design: A test group and a control group are randomly selected; the criteria variables are measured in both groups; the test group receives the program treatment while the control group does not; and the criteria variables are measured again. This design can be modified to test several alternative program strategies at once by adding a randomly chosen test group for each alternative strategy.

The true experimental design is the most difficult of the evaluation programs, but the validity of the results is the highest attainable (Acland 1979). The design is most appropriate for programs that are intentionally experimental in nature (demonstration or pilot programs); that are implemented in small-scale, flexible units (e.g., treatment or service programs); and/or that involve high risks and unacceptable uncertainty about potential benefits and costs. Generally, such designs are not appropriate to large-scale capital investments except when statistical control jurisdictions can be identified. On the other hand, the accuracy of experimental designs is balanced by the large number of problems that potentially can degrade the quality of the results: the results of a small-scale experiment may not reflect what might happen when the program is implemented on a large scale; self-selected subjects may not represent the behavior of the general population; and it may be inappropriate to withhold the program treatment from one group or to provide it only on a temporary basis (Rivlin 1971). The costs and required skills of experimental designs are substantial because of the effort required to plan and manage the experiment and to collect and analyze the data.

3.2.3 Naturalistic Inquiry

The above discussion of evaluation design, measurement, and data collection emphasizes the use of quantitative techniques. Evaluators increasingly are recognizing, however, the utility and even necessity of some of the research methods used in anthropology, political science, sociology, and social psychology. These methods sometimes are called "qualitative," but David Kline argues that "naturalistic inquiry" is more appropriate (Kline 1978). Whatever its name, the main concern is "to obtain more valid and complete data about the phenomenon being studied than can be obtained through the reactive and reductive techniques of surveys and experiments" (Kline 1978: VII-17). Naturalistic inquiry, thus, plays a valuable role in evaluation research. It can develop knowledge about

the actual behavior of public programs that cannot be quantitatively calibrated, such as the social values underlying a move to decentralized power sources or the quality-of-life estimates. In combination with the experimental or quasi-experimental methods mentioned above, naturalistic inquiry may be valuable in illuminating the reasons why public programs perform the way they do.

There has been a reluctance among many evaluation specialists to employ qualitative measures, usually out of concern for internal validity of data. Everett Rogers admits this has been one of his concerns but that he now views the neglect of qualitative methods as wrong:

[Too] many times we have defined evaluation just to mean quantitative evaluation. . . . [Yet] we all agree that the quantitative evaluation has to be embedded and placed within a larger context. We often ignore that; . . . but to most people evaluation has come to mean only the quantitative evaluation, and that is a terrible mistake.

Naturalistic inquiry is prone to internal validity and reliability problems, but in the hands of a careful researcher, it can suggest many key findings, such as the role that institutional biases have in shaping programs and how social and individual values can be included.

3.2.4 Comparison of Evaluation Designs

The basic evaluation designs may be selected according to several practical characteristics: cost, personnel requirements, validity, time, data collection requirements, program scale, management involvement, and impact on decision making.

The cost of the basic designs generally increases in the order presented: case study, one-group or static comparison groups, time series, nonequivalent comparison groups, and true experimental control groups. The technical skills required to perform the evaluation also increase in the same order, as well as the time needed to conduct the study and the extent and rigor of data collection.

Very large-scale programs usually preclude the use of rigorous experimental designs because of the difficulty of finding equivalent control groups. For example, the federal income tax credit for solar energy equipment cannot be evaluated by an experimental design because the program applies to the entire national tax-paying population. A time-series design may be appropriate for some such large-scale programs, but, in the case of solar tax credits, the small baseline and recent rapid growth in solar heating use (stimulated in some measure by federal demonstration programs) would make the actual effect of the federal tax credit difficult to discern. This problem indicates the difficulty of evaluating the effectiveness of any large-scale public program that has not been tested first on a pilot basis.

Management involvement in the planning and execution of the evaluation study is greatest with experimental designs, where the staff has to maintain consistency in the performance of the program actions and also may be responsible for the data collection.

The effect of the evaluation on decision making is a product of the validity, the relevance, the receptivity, and the timing of the evaluation results. Experimental designs provide the most valid information, but their narrow focus may not be pertinent to the

decision maker's needs. Also, the long lead time required to plan, conduct, and analyze experiments may mean that the results come too late to be useful. Pre-experimental designs may have an impact on policy because they can be done quickly enough to respond to immediate interests and are flexible enough to be relevant to a range of concerns, but their validity is highly case dependent. The quasi-experimental designs represent a balance between expedience and rigor.

Another means to assess the relative merits of the basic evaluation designs is to apply them to different types of programs. Technology R&D programs notably are difficult to evaluate, although case studies may generate some interesting information. Quasi-experimental or experimental studies may be useful in examining generic R&D processes but might not indicate the effectiveness of specific R&D programs. Commercialization programs are suited to both quasi-experimental and experimental approaches to evaluation. Demonstration programs are intrinsically experimental in nature and can utilize experimental designs. Information, communication, education, and financial programs may be evaluated by either experimental or quasi-experimental methods, depending on what is most appropriate for particular programmatic needs. Legal restraints and scaling problems usually make experimental evaluation of regulatory programs infeasible. Non-equivalent comparison group designs may be most practical for evaluating state and local regulatory programs. Public works programs involving large-scale capital investments may be evaluated with comparison group and time series methods.

3.3 MEASUREMENT

A necessary feature of any program evaluation is some form of measurement of the program's effects. This section briefly reviews the concepts of what measurement is and the basic criteria of measurement quality—reliability and validity (Kline 1978: X-7 through X-12).

Reliability is associated with consistency of results and is based on stability, equivalence, and internal consistency. Stability is the tendency of a given test (measurement technique) to produce the same result when repeated under the same conditions. Equivalence is the tendency of different forms of the same test administered simultaneously to produce the same result. Internal consistency is the capacity of different parts or items of the same test (split-half or odd-even) to measure consistently when simultaneously administered.

Validity is the relationship between the measurement and the actual property of the thing measured. There are four major tests of validity: content, concurrent, predictive, and constructive. Content validity depends on how well the test represents the content of its subject. For example, does an examination in American history really measure knowledge of American history? Concurrent validity asks how well the test results conform to some other accepted performance standards; e.g., does a dynamometer test measure an automobile's fuel economy under actual driving conditions? Predictive validity is a function of how well the test results predict future performance; e.g., do scores on the Law School Admission Test predict student performance in law school? Constructive validity examines how well the test results conform with the predictions of theory. An illustration: if current motivational theory predicts that people who live in glass houses are unlikely to throw stones, does a survey of residents of various types of houses indicate a high correlation between glass house dwelling and absence of stone-throwing behavior?

The above estimates deal with what is known generally as internal validity (i.e., the authenticity of measurement in relation to the thing measured). Another important concern of evaluation research is the generalizability of the results of an evaluation study to other situations or external validity. Threats to external validity are just those factors that make the evaluated program incomparable to other, similar activities. Ways to increase external validity include evaluating the program under as natural and unobtrusive conditions as possible, evaluating activities that are typical of program activities elsewhere, evaluating a wide variety of program activities or projects to measure the effects of a diversity of conditions that may exist in other situations. Increased experimental control tends to decrease external validity as it increases internal validity (Rutman 1977: 35-36).

3.4 DATA CONSIDERATIONS

Data collection and definition are necessary for the measurement of program results. For convenience sake, one can identify five basic data collection procedures: interviewing; self-reporting; observing; unobtrusive measuring; and mechanical or electronic recording. The first three are known as direct or obtrusive procedures because the subject is aware of their application.

Interviewing involves the direct questioning of subjects from whom information is sought in either scheduled standardized, unscheduled standardized, or unstructured formats. In a standardized interview, the interviewers ask every subject the same set of questions. In a scheduled interview, the questions are written out in advance and the interviewers read the questions in exactly the same way to every subject. In an unscheduled standardized interview, the topics to be covered are established in advance, but the interviewer has flexibility in how the questions are posed and may ask the subject to clarify or expand on answers. In an unstructured interview, only the general theme or topic is established in advance; the interviewer and subject pursue issues as they develop.

Self-reporting permits the research subjects to provide information without an interviewer present. Self-reporting instruments include questionnaires, report forms, and standardized tests. Questions and test items may be structured many different ways to provide information suitable to a variety of scales and measures.

Observing is a third data collection procedure. The researcher observes subject behavior or activity and measures it either by counting or judgment. For example, evaluating teacher performance using direct observation might involve counting the frequency of a set of classroom events such as the number of students called on in a period. Observation also might use subjective judgment reported on some pre-established scale.

The three direct methods of data collection have the disadvantage that all might conceivably affect the behavior they intend to measure, the so-called "Hawthorne" effect. Unobtrusive measuring sometimes is used to control the biasing effects of direct procedures. Census information, reports, and trade statistics are common forms of unobtrusive procedures. An anonymous observation, such as traffic flow, is another example. Mechanical or electronic recording may be used as either obtrusive or unobtrusive procedures.

The determination of what data should be collected and analyzed is more important than the means of data collection. For a number of reasons (e.g., costs, expediency), evaluations often are based upon readily available data that already have been collected. In

many such cases, insufficient attention is paid to whether the data are suitable for the evaluation task; convenience takes precedence over rigor. At times, this situation cannot be avoided, but, when it occurs, it should be recognized and the appropriate reservations inserted. For evaluations in which time and resource constraints are less demanding, data generation should be part of the evaluation design and considerable effort expended towards making certain that the data fit the evaluation requirements. In these cases, care need be taken that the data definition task is solidly grounded in a theoretical framework or risk the pell-mell and expensive collection of data that later prove of dubious value (de Neufville 1978: 172). The history of the social indicators effort illustrates the danger of collecting data that lack a theoretical underpinning (Sheldon and Freeman 1970; Bauer, ed. 1968). Hard questions—such as what is the objective of this evaluation? What theories are applicable and why? What data reflect the theory? And, how do these data fit into the evaluation design?—must be answered before beginning the data collection.

3.5 EVALUATION DESIGN AND WORK PLAN

Evaluation studies require careful, methodical planning. "The result of unplanned evaluation programs has been a proliferation of noncomparable, poorly designed studies with little or no policy relevance" (Wholey et al. 1976: 46). Ideally, evaluation planning begins before the program is initiated and extends to a detailed work plan for executing a specific evaluation study. In their study of federal evaluation policy, Wholey et al. (1976) made several recommendations for developing evaluation work plans. Among others, they suggested that approved evaluation plans should be required at the beginning of each fiscal year as a condition for authority to spend evaluation funds; that evaluation work plans should cover a two- to three-year period and should define (1) specific questions to be answered on the effectiveness of selected programs and the dates by which answers are needed, (2) estimated cost of each study and resources available for the studies, (3) agency priorities among studies listed, and (4) important questions that will not be addressed; and that an assessment of methodological feasibility be conducted possibly in terms of the probable value of findings, comparability with related evaluations, and usefulness relative to other proposed studies.

Chemlinsky's report on evaluation by federal agencies recommended that evaluators be involved in program planning activities and that local practitioners be included in the planning, design, and development stages of programs to ensure that evaluation criteria reflect the interests of all the participants. Similarly, cognizant congressional staff should be involved in planning for major program evaluations to encourage clarification of legislative objectives and increase congressional use of evaluation findings (Chemlinsky 1977: 49).

Nay and Kay assert that the evaluation design for a particular study needs to be translated into a detailed work plan that specifies the boundaries of the system to be evaluated; the testable model, which includes the client's understanding of the system to be evaluated and what it expects to learn from the evaluation; the evaluable measurement model, which indicates the relevant variables and how they will be analyzed; the analysis plan; the data collection and measurement plans; the information to be produced, including the form and media for communicating the final results of the evaluation; and the budget (Nay and Kay 1978: 102-123; also see Walling et al. 1978).

3.6 COMMUNICATION OF EVALUATION RESULTS

The key requirement for communicating evaluation study results is the written report, which should cover the evaluation, background, statement of rationale, purpose and objectives, program description, and methodology. The key sections are, of course, the analysis of results and the policy recommendations that can be deduced (Kline 1978: X-18). Agencies involved in the activities being evaluated should be permitted—indeed, encouraged—to review evaluation findings prior to publication and should be given reasonable opportunity to respond.

The communication of evaluation results too often ends with submission of a lackluster, written submission to the client. This may be one cause for the uncertain reputation of evaluation results. Since the essential purpose of evaluation is to inform program managers, research products that are not communicated effectively should not be considered complete. Written reports should be composed and edited carefully to be as understandable and useful to the client as possible. In addition to written reports, oral briefings and other presentation formats can be used to increase the dissemination and impact of the evaluation (Meltsner 1979).

Evaluations usually are prepared with the intention that they are public documents and free from agency censorship. With some special exceptions for national security, evaluation reports cannot be legitimately sequestered. Still, public agencies should be aware that even draft evaluation reports rarely can be kept confidential. While evaluators must guard their independence and prior restraint on the publication of evaluation results, they also must protect the legitimate privacy or confidentiality of some sources of information. The ethical standards for evaluators should be viewed as stringently as those for journalists, lawyers, or physicians.

In general, the purpose of program evaluation is to evaluate programs, not people. The information an evaluator gathers about individuals usually should be considered privileged unless there is an overriding public interest in its disclosure. Promises of confidentiality should be made only when necessary and, having been made, should be honored. On the other hand, public programs are public business with few exceptions. Protecting government officials from public embarrassment as a result of their official performance is not part of the evaluator's job. The evaluator's role is to assist his client in obtaining and analyzing accurate, fair, and reliable information about the consequences and value of government activities. If officials are embarrassed by such information, the fault is more theirs than the evaluator's.

Important evaluation findings should be communicated to as wide an audience as can benefit from them, which often lies beyond the domain of the immediate sponsor of the research. Indeed, open publication of evaluations is often a contractual requirement. Findings that are relevant to the public management or to social science literature should be published in appropriate journals. Evaluation results that are germane to critical issues of public importance should be communicated to popular media. For example, the press reported a number of evaluations of the President's Domestic Policy Review on Solar Energy.

3.7 META-EVALUATION: EVALUATING EVALUATION

Evaluation itself should be no less subject to critical evaluation than any other publicly funded activity, for it is consciously designed to address public problems. Several important desiderata are used to assess evaluation results, including internal and external

validity, evaluation data and methodology, and evaluation relevance to and influence upon program management and public policy (which includes the effectiveness of the communication of its findings).

The performance standards of professional evaluators (as implied by journal publications) may not be the best guide to the real quality of evaluation results if the ultimate objective is to influence public policy. In that case, the final judges of the evaluation research are the clients and consumers of the work itself. In some cases, the results could be immediate, as was the case of the critical evaluations of performance contracting in education, in which program managers moved quickly to correct the abuses revealed by Carpenter and Hall (1971). In other cases, the results may be longer in coming if, as Weiss (1977) suggests, policy makers primarily use evaluation studies as a general source of good ideas and orientations. The methodological and policy debates engendered by documents, such as the Coleman report on school desegregation (Coleman et al. 1966), can only bode well for all evaluation efforts (see Mosteller and Moynihan, eds. 1972), both from policy and meta-evaluation perspectives.

SECTION 4.0

ORGANIZATION AND MANAGEMENT OF PROGRAM EVALUATION

This section reviews the basic requirements for the organization and management of program evaluation activities. The major issues considered are the evaluation personnel, sponsorship and funding, use of internal versus external evaluators, timing of evaluation, and confidentiality. The emphasis is on the essential requirements for organizing and conducting formal evaluation as well as many of the practical problems of evaluation.

4.1 EVALUATION PERSONNEL

Personnel are the key element to evaluation. Evaluation is not a process that can be delegated to computers or routine procedures. The core of any useful evaluation is human judgment of process, effect, and value.

Despite the crucial importance of the personal qualifications of effective evaluators, the evaluation literature pays scant attention to qualifications. Therefore, a series of interviews with evaluation experts focused particular attention on the necessary characteristics of evaluation personnel, specifically the professional skills and human resources they would seek if they were establishing a solar energy program evaluation group and the necessary characteristics for an evaluation manager. Although the experts' specific prescriptions varied, there were numerous points of consensus. First, the interviewees indicated that an effective evaluation group needs to be multidisciplinary, representing a balanced diversity of training and experience. Some emphasized that by multidisciplinary they do not mean generalists but a group of experts who are thoroughly trained in their particular discipline yet able to communicate effectively with each other. Economics, political science, and sociology were the disciplines most often referenced; survey research, quantitative analyses, and a familiarity with technology were the skills cited as being especially useful.

Although those interviewed saw a need for a variety of disciplinary and professional specialists, most placed priority on a more generic set of professional skills, combining intelligence, an understanding of causal relationships, skills in basic evaluation methodologies, and an appreciation of organizational and managerial behavior, especially how they affect the programs. Most of the experts interviewed stated that the important qualifications for effective evaluators are based more on innate ability and proven experience than on formal credentials. Victor Berlin of the National Bureau of Standards summarized what many of the experts expressed in terms of evaluation skills:

Do you know how to go out into the real world and deal with people and find out what their needs are, find out very carefully how the systems work, how they interact? Can you convey back to people and communicate with people who are not in your discipline . . . exactly how things work and what the limitations are? Do you know how to build a coalition around a project, build a constituency? Do you know how to work through an organization, trace through an organization or an agency or an industry and find out how it works, find out how the functional elements of it work?

It's someone who has probably worked in and had engineering, psychology, sociology, political science, and systems work. . . . They have to be smart and they have to be open-minded. They have to be entrepreneurial, risk

takers. Those are the characteristics that are important. They have to be able to get into an area fairly quickly, read up on it, understand it. . . . People who know how to write well—they've got to write really well. They have to be good at communicating. They have to be careful, they have to be politically sensitive. I can list the qualities; you find the person. If they have the qualities, fine. It doesn't matter what their background is. I think the person is more important than the label.

The qualifications mentioned for an evaluation manager of solar technology programs emphasized the demonstrated capability to organize and manage; good evaluators are not automatically good managers. They should possess excellent technical skills and a thorough knowledge of evaluation theory and practice. Finally, they should be able to frame issues in terms of public policy problems and solutions. There was a general consensus that the knowledge of specific solar technologies was less important than the above qualities; knowledge of solar technology was viewed as an expertise that could be readily acquired or derived from outside sources. One expert went so far as to argue that solar energy technologists should be excluded from a primary role in solar program evaluation for two reasons. First, they are likely to have established associations with particular technologies which might bias their assessments. Second, program evaluation tends to be a "high mortality" profession. Critical evaluations can be politically embarrassing, which could be damaging to technical specialists who might be more vulnerable to negative career impacts from controversial evaluations rather than professional program evaluators who could more easily shift their skills from solar energy to other program areas if necessary. Given that technical skills are of some importance, these negative effects should be minimized by locating such individuals in professional evaluation groups, such as the GAO or university-based organizations.

4.2 SPONSORSHIP

We can identify five basic arrangements for the sponsorship and financial support of evaluation studies. First, a given agency can provide sole support for evaluation activities within its purview. Alternatively, the agency could provide support for administration and some continuing features of program evaluation, but specific evaluation projects would be funded by external sources. Third, the agency may participate with other agencies in a research consortium, with shared support. Fourth, the agency might contract to have the evaluation performed by an independent research organization. Last, the evaluation group can be made strictly accountable to an independent, outside organization (Twain 1975: 33-34). The money for the evaluations can come from the normal budget requests, special ad hoc appropriations (e.g., a presidential task force), discretionary funds, a special agency "tax" designed to support program evaluation, or an outside organization (e.g., NSF).

No single arrangement for organizational responsibility is suitable for all situations, but as a general rule of thumb, "In most cases [operating agency heads] should assign overall responsibility for conducting the evaluation to a unit at a higher organizational level than the program being evaluated" (Hatry et al. 1973: 114).

The total budget for program evaluation is difficult to prescribe; it may range from 1% to 15% of total program funds, depending on the requirements for evaluation and the nature of the particular program.

4.3 USE OF INTERNAL VS. EXTERNAL EVALUATORS

An important practical issue in the organization of evaluation studies is the use of internal evaluators versus evaluators from outside the agency or program being reviewed (Hatry et al. 1973: 117-119). External evaluators may be from either universities or private research organizations or they may be government employees from an independent agency charged with evaluating public programs, such as the GAO or the Congressional Budget Office. External consultants, especially those outside the government, may serve as supplementary labor for tasks that may not warrant a full-time staff person, such as designing and conducting a survey instrument. An external consultant may offer training and criticism to an internal evaluation staff on evaluation studies. Finally, the external consultant may assume the full responsibility to design and execute program evaluations.

The use of external evaluators has both advantages and disadvantages, the key criteria being efficiency, effectiveness, and credibility. In regard to efficiency, external evaluators may have special skills, resources, and experiences that may enable them to produce useful evaluation results with less investment of time and money than an in-house review. Especially where evaluation contracts are awarded through competitive bids, more efficient evaluation services may be obtained. However, sometimes the costs of using an external evaluator may be greater than an internal evaluation because the outsider needs additional time to become familiar with the program and the organization. This is particularly relevant for evaluations where time horizons are immediate.

The use of external evaluators may improve the effectiveness of evaluation studies in several ways. The skills, resources, and experience of the external evaluator may be superior to what is available within the agency. The naivete of the external evaluator regarding the program actually may be an advantage when the evaluation requires an impartial observer in order to gain credence. In addition, the independence or neutrality of the external evaluator may reduce possible resistance of program personnel to assisting the evaluation. Also, external evaluators may be able to communicate results to program officials more directly and more frankly than could internal staff.

Alternatively, reliance on an external evaluator might reduce the effectiveness of the evaluation in some ways. The external evaluator may not have an accurate perspective on the program and decision-making processes. As an outsider, the external evaluator might be perceived as more threatening to the internal group and, hence, could face increased resistance to the evaluation from program personnel. And the lack of a continuing relationship of external evaluators with program officials could mean that the evaluators will not be in a position to promote the use of evaluation results in subsequent decision making.

Credibility or relative objectivity may be the issue most commonly associated with external evaluators. Credibility goes beyond objectivity. Unobjective evaluators stand to lose their credibility, but relatively unbiased, candid internal evaluators may lack credibility simply because of the appearance of their relationship with those in positions of authority. The "where you stand depends on where you sit" syndrome is hardly clear-cut. Private consultants often are concerned about their future work with the agency and are so hesitant to be disruptive that they, in effect, can be "bought"; government evaluation offices and agencies are likely to be rewarded for critical evaluations. Thus, both groups' credibilities and priorities are questionable. In theory, the external evaluator starts out with a greater degree of independence and, hence, credibility; but credibility ultimately becomes a product of the reputation and record of particular individuals and organizations, whether internal or external.

4.4 TIMING OF EVALUATION

The timing of evaluation in relation to government decision processes is one of the most important concerns of evaluation management. Ideally, evaluation should be planned to respond not just to the issues of the moment but to future questions, problems, and developments.

One of the serious problems in the timing of evaluation is that the effective demand for evaluation in some program areas is virtually nonexistent until a crisis occurs. In such instances, the planning and execution necessary for thorough evaluation cannot be performed and the imperative for any evaluation takes precedence, almost certainly resulting in an unsatisfactory—if not erroneous—evaluation. At such a critical juncture, poor program evaluations offer too little too late to save the program and do nothing to enhance the relevance and credibility of evaluation activities themselves. This was certainly the case when the Nixon administration demanded an evaluation of the New Jersey income maintenance experiment before it was completed (see Rivlin 1971). Thus, it is both methodologically preferable and practically politic for program managers to initiate evaluation plans and procedures at the inception of their programs as preparation for the inevitable day when they are called upon to explain their activities and justify their continued budget.

The recently completed evaluation of the Energy Extension Service (EES) pilot program may be a successful example of an anticipatory strategy for evaluation. The EES evaluation was initiated by the program managers in anticipation of a future demand for a critical appraisal of the pilot phase of the EES program before the program would be given full-scale funding and established for all states. This decision was made at a time when Congress was not only enthusiastic about the pilot program but requested budget proposals for a full-scale EES program prior to the evaluation of the pilot phase. More recently though, Congress was considerably more critical and questioned whether the EES program should be continued at all. Whatever the technical results of the EES evaluation, it already appears to be prescient in terms of timing and foresight. This experience could be an important model for solar energy program evaluation, assuming that the EES program is sufficiently isomorphic to be relevant to other solar technology programs.

Lastly, an agency should resist the premature evaluation of a new program for fear that it might stifle creativity and reduce the number of expected, and even necessary, mistakes that one associates with any new and uncertain endeavor. In essence, a new program merits the right to succeed before being evaluated as a failure (deLeon 1978).

4.5 CONFIDENTIALITY

Confidentiality in the performance of program evaluation is an issue not only guided by professional standards but also mandated by law. At the federal level, the Freedom of Information Act and the Privacy Act govern the gathering, use, and reporting of information in evaluation studies. At the state level, similar laws exist that may be more or less stringent in many cases than the federal codes.

The evaluator's responsibilities combine some aspects of the roles of journalists, physicians, and lawyers. Evaluators have the professional obligation to withhold information under certain circumstances as well as to disclose information under other sets of circumstances. At present, evaluators have not had the time nor incentive to codify a body of experiences and distill an ethical standard, although some attempts have been made

(see Operations Research 1971 for a critical assessment of a policy debate). The creation of professional ethos, standards, behavior, expectations, and sanctions will be a future major issue that could have an effect on solar energy evaluation. This existing ambiguity lends weight to a recurring theme in this report, that evaluation—for all the time and money expended and reports issued—is still inchoate and rapidly evolving. To demand or expect hard rules of behavior or procedure is currently impractical and impolitic.

4.6 PRACTICAL EVALUATION PROBLEMS

Thus far, this report has concerned evaluation desiderata that, although pressing, were somewhat technical and removed from operational difficulties. There is another class of evaluation difficulties—the everyday realities of evaluation that go beyond textbook prescriptions of classifications, measures, and procedures. Evaluation requires direct involvement with real people in functioning organizations, each with its own frailties, each may appear to be doing unproductive tasks for irrational reasons. In this mundane context, evaluation attempts include formulating rational judgments based on criteria of quality, equity, and effectiveness. This section discusses some of the practical aspects.

4.6.1 Needs and Purposes

About a decade ago, the relationship between universities and their local communities became a typical concern in higher education. Manifesting this, the administrators of a psychiatric college and teaching hospital in Canada decided to engage in a more active contribution to the welfare of the surrounding community. A study was made to determine what new service the hospital could provide to the community. Upon examination, it was learned that the community had the highest suicide rate in all of Canada, a serious community problem that the hospital seemed ideally qualified to address. Based on a program that had been tried elsewhere, the hospital decided to organize a 24-hour "crisis clinic" where potential suicide victims could receive emergency treatment to resolve the immediate crisis and where they could be directed into some longer term therapy.

A year after the crisis clinic had been in operation, new public health statistics showed that the city's suicide rate had increased significantly. The administrators of the crisis clinic treated this as evidence that some new factor had been introduced to exacerbate the already serious suicide problem. Because prevention of suicide was their goal, the administrators investigated the cause of the reported increase in suicides, expecting that their program might have to be expanded and otherwise improved. The study disclosed that the primary cause for the increase in the number of reported suicides was the crisis clinic itself. The clinic quickly had become recognized throughout the community as a valuable institution with a staff known to assist those who came to the clinic in resolving whatever their problems were. However, like any institution, the crisis clinic had regulations, one being that the potential patient had to show evidence of suicidal tendencies to be eligible for the clinic. Many of the city's disadvantaged population, therefore, simulated suicidal acts—slashing their wrists or taking drug overdoses—as a means of qualifying for help from the crisis clinic. People who could not pay their electric bills, people who were ill and could not get medical care, people whose dogs had run away—people with any kind of problem who could not get help from other public agencies—would fake a suicide attempt so that they could be admitted to the crisis clinic. Unfortunately, many of these simulated suicide acts were inadvertently successful; hence, the increase in the rate of suicide in the city. Following the investigation, the crisis clinic was closed.

This episode presents four observations that are germane to program evaluation. First, good intentions are no guarantee of good results. Merely because a program is designed to serve the public interest or because program staff think it is serving the public interest are insufficient reasons to assume that the program actually is serving the public interest. Objectives need to be defined against which program results can be assessed, not wished.

A second observation is that success can be illusionary. Programs may be enormously successful in meeting their immediate operational objectives but still be counterproductive in their effect on larger goals. Within a narrow context (e.g., patients treated), the crisis clinic program was a great success; but within the context of the entire community, the program was a failure.

The third possible lesson for evaluators is that things are not always what they seem. The crisis clinic was an institution dedicated to the prevention of suicide, but it performed a variety of social functions only indirectly related to suicide. Not only were the staff's perceptions of the clinic different from the clients' perceptions of the clinic, but the staff's perceptions of the clients' perceptions were erroneous in a way that proved literally fatal. A social system is a complex, dynamic system that is capable of creating new and different problems. Over time, the system strives to attain some sort of dynamic equilibrium. To treat a single problem, such as suicide, without a full appreciation of the contextual intricacies and interrelationships of the system is a risk to the delicate systemic equilibrium and possibly results in worse conditions than initially encountered. All these lead to a last observation that there is no substitute for an accurate assessment. The basic goal of program evaluation—whether formal or informal—is to determine what is going on, why, and what can be done to correct possible problems. Models, interviews, surveys, and data all may aid understanding of what a program is doing but, if they do not lead ultimately to an accurate accounting, they are merely empty techniques and may be even worse if they obscure what is really happening. In the case of the crisis clinic, the cure was in fact exacerbating the disease.

4.7 PROGRAM EVALUATION IN THE POLITICAL PROCESS

One of the major obstacles to effectiveness of the program evaluation in affecting policy decisions is the lack of integration of program evaluation with other elements of the policy process. Evaluation results rarely are used to design new programs. Many programs never are evaluated because there is no demand to do so; hence, there is no opportunity to learn from them. Other programs escape evaluation because they literally are set in concrete (e.g., super highways and irrigation projects) so that their faults would be nearly impossible to correct.

4.7.1 Reasons Why Evaluations Are Not Done

Although the number of evaluations is large and increasing, the existing amount of evaluation work still is slight compared to the total expense invested in specific public policies and programs. There are many reasons why evaluations are not done more comprehensively and more frequently. The most obvious is that the long-term benefits of evaluations to program managers and decision makers have been less persuasive than their short-term costs in terms of scarce resources. The value of evaluations has not been demonstrated, especially to the policy makers whose perspective rarely extends beyond the next fiscal year. This opinion is not altogether unsubstantiated, especially when the

quality of past evaluation work in terms of validity, relevance, timeliness, and results, often has been substandard. Program evaluation is undercut by program managers who are prone to argue that "the value of my program cannot be measured." Again, they are not necessarily being obdurate; what can be measured often is limited and may not be relevant to pressing decisions. Program evaluations can be and often are controversial. Evaluations can be threatening to many vested interests and organizations, a condition which makes evaluation that much more difficult. Finally, it must be realized that evaluation under the best of circumstances is a difficult art. Government agencies often lack personnel who are skilled in formal evaluation techniques and analysis. Even when staff have some relevant training and experience, they may not have the specialized skills required to carry out a full-scale evaluation. Under these constraints, it is hardly surprising that evaluation is not more widely accepted and practiced (adapted from Hatry et al. 1973: 20-21).

4.7.2 The Psychological Obstacles to Evaluation

Evaluations interact strongly with human perceptions and motivations. Sensitive evaluators must understand the significance of human behavior at the individual level, at the organizational, social, and political levels, and even their own personal behavioral patterns.

Evaluation often is perceived as a critical, destructive procedure and, therefore, is resisted by the organization and personnel being subjected to evaluation. The professional reluctance on the part of the staff is reinforced by the societal condemnation of failure. If the staff views evaluation as leading to the detection—indeed, the certification—of their personal and programmatic shortcomings, then the evaluator is forced to work in a psychologically hostile environment. This is the psychological reality in which many evaluations are conducted, a condition that makes a balanced evaluation more difficult than would be the case if the evaluation were perceived in a more beneficial or at least benign mode (see Campbell 1969 or Campbell in Caro, ed. 1971).

Another possible basis for the psychological resistance to evaluation is the need to maintain a set of consistent intentions or orientations or, put another way, to avoid cognitive dissonance (Festinger 1957). Simply put, people are more likely to remain in familiar behavior patterns, to do what they do (or know) best. If evaluation results threaten to undermine these patterns, to force them to confront dissonant information, then they will go to great lengths to avoid such a conflict even if it requires a conscious distortion of reality (see Steinbruner 1974). Organizations are no less willing to distort or subvert reality to fit their own particular needs (Janis and Mann 1977).

The enumeration of psychological barriers to evaluation does not demonstrate or even suggest that they are foreboding obstacles to all evaluations. Rather, they illustrate the kinds of human motivation problems that evaluators need to consider when formulating and conducting program evaluations. Fundamentally, program evaluation is an intensely humanistic enterprise, much more complex than simple measurement, collection, and analysis of data.

4.7.3 Evaluation for the "Wrong" Reasons

All the rhetoric about the important needs for evaluation found in literature on the subject may obscure the fact that evaluation occasionally is done for reasons that stand in

opposition to its nominal purpose. Edward Suchman identified five common reasons—what he called "covert purposes"—for doing program evaluations, which have little to do with evaluation per se.

Suchman refers to the first of these as "eyewash" or the deliberate focus on the surface appearance of a program to make it appear better than it is, while the second is the proverbial "whitewash" or an attempt to cover up program failures. Other misuses of evaluation include the "submarine" (i.e., the political use of research to destroy a program), "posture" (evaluation research undertaken only because it was a precondition), and "postponement" or using evaluation to delay action (Suchman 1972: 81).

The enumeration of these excuses for evaluation is not meant to suggest they are without political and even programmatic merit. Often, a postponement can serve everybody's best interests by defusing a volatile issue. The point here is not the relative merits of such tactics, rather that the tactics should not be used as reasons to initiate formal evaluations because their underlying purposes are at odds with the objectives of evaluation.

4.7.4 Ambiguity of Political Decision Making

Pluralistic decision making is inevitably rooted in compromise. Inconsistent political interests often combine to create a particular government program representing diverse goals. The result is that the objectives found in legislative and executive mandates for government programs can be ambiguous and occasionally contradictory, thus providing the evaluator scarce measures of understanding the program's objectives and few measures against which the program can be evaluated. From the evaluator's point-of-view, the problem is:

[a] lack of clarity and consensus on the objectives of government programs. Too often objectives are utopian, and they are usually vague. In many instances, programs are stated in such convoluted prose that there is no way of telling except for intuitive feeling whether their objectives are accomplished or not. The program authorization process itself involves compromise among opposing positions. These are reflected in ambiguous program objectives. (Chelmsky 1977: 34)

The ambiguity of legislative goals is an integral feature of the legislative process, an inevitable by-product of compromise and coalition-building which are usually necessary for the passage of important legislation. The solution to the problem of legislative vagueness will not be found in the political process itself, although some efforts are underway to persuade legislators to define program goals more explicitly and even to designate some evaluation metrics. This was the thrust of much testimony offered during the 1978 hearings on the federal "sunset" act (U.S. Senate Subcommittee 1977). There is no easy solution to this problem of ambiguous rhetoric given the pluralistic nature of the political system. The only conclusion one can draw is that this condition increases the difficulties faced by the program evaluator.

4.8 THE MANAGEMENT ENVIRONMENT OF EVALAUTION

In practice, the management environment is replete with institutional problems that must be considered for program evaluation. For instance, although evaluators and program officials need to cooperate in conducting an evaluation, a critical tension can easily exist between them over the control of the evaluation process. Evaluators tend to resent

administrative interference in the design and execution of the evaluation, while administrators justifiably worry about how the evaluators will disturb the operations of their program, not to mention what evaluators might uncover. On a lower level, program staff are likely to protect themselves and the program from potential criticism by censoring, sequestering, or seriously biasing information that evaluators request, thereby making the already difficult evaluation effort even harder.

Data collection requirements may appear to program staff simply to be more unnecessary paperwork, with the result that information may be reported inaccurately, or sometimes not at all, as a low priority item unless it jibes with the institutional perspective. Likewise, program staff undoubtedly will place their service obligations ahead of their evaluation requirements. The evaluation could be subverted further should the staff decide to change the program activities without informing the evaluator. Different work schedules and methods may result in friction between program staff and evaluators. Evaluators may keep irregular hours while the staff works actively on a fixed schedule. This condition could be aggravated by status conflicts that could arise between evaluators and program staff. Evaluators and staff are likely to have very different academic and professional backgrounds. Either may be guilty of parochialism which may easily lead to a lack of respect for the professionalism of the other. Finally, the publication of evaluation results may lead to conflicts between evaluators and program officials over content and credit (Caro, ed. 1971: 10-12).

Again, these issues do not permit ready resolution because of the very nature of evaluation exercises and organizations. At the least, they can be recognized and explicitly confronted. At best, they can be ameliorated by the close and active cooperation of the program manager and the evaluation group leader. If these two can agree, then their cooperation might extend to members of their respective organizations.

4.9 THE USE OF EVALUATION RESULTS

A major concern of those skilled in evaluation is the difficulty often experienced in bringing the results of the evaluation to the attention of decision makers. The results of much of the evaluation work that has been done have not been useful and understandably have not been used. While there was general recognition that this condition exists, there is disagreement over whether it is deserved. John Ware is skeptical about the utility of many evaluation studies:

I just find it very discouraging. I often hear people like heads of [evaluation] departments . . . being asked, "Give me three examples of something that you've done in the last ten years that has impacted on some policy decision." Boy, are they hard-pressed to come up with a trivial example, and somebody in the audience is going to raise issue even with that one, as focused as it might be.

David Armor, among others, argues to the contrary that there are numerous examples of valuable evaluation studies that have had a significant impact. Launor Carter concurred and further asserted that the utility of evaluation often goes beyond its immediate application to the particular program evaluated, that evaluation can be used in three indirect ways which may be more valuable than the direct use by program managers. First, evaluation may stimulate public debate about important program issues which eventually affects the formulation of new policies and programs. Second, evaluation research may generate information that decision makers other than the immediate client may need at

a later time or for a different reason. Finally, extensive evaluation experience in a certain program area provides the evaluator with valuable experience on the design and performance of such programs.

Whatever disagreement may exist, the overall utility of evaluation, and the underlying reasons, it appears that the results of many evaluation studies are not translated into programmatic actions or concrete policy recommendations, largely due to four causes. First, and most obvious, the results of the evaluation could be erroneous or misleading because of methodological flaws in the evaluation design, execution, or analysis. At times, the mistakes can be flagrant and easy to identify, such as the use of an inappropriate model or a faulty data base; other times, the mistakes could be much more subtle and open to serious debate if discovered, such as over the proper uses of multiple regression models and data specifications that characterized the methodological debate over the Coleman report (see Mosteller and Moynihan, eds. 1972). Of course, an evaluation can be criticized for its methodology when, in fact, the germ of contention is its political implications, as in the various evaluations of negative income tax policies (Moynihan 1973; Rivlin 1971) and OEO programs (Levine 1970).

A second reason why evaluations may not be translated into program activities is that they were never intended to be; in Suchman's phrase, they are evaluations for "covert purposes." Some evaluations are not utilized because their results were not timely; that is, they came after important decisions regarding the program had been made. Although not a program evaluation, strictly speaking, the SERI photovoltaic venture analysis (Costello et al. 1978) came after Congress had decided to fund a large-scale production program for photovoltaics; thus, the SERI criticism of the market-push strategy came too late to be useful in designing the program.

The third reason why evaluation results might not be used is that they do not meet the requirements of the decision maker or program manager, regardless of the quality or timeliness. For example, an evaluation that describes only the program effects but not the processes which produced them provides no link between actions and effects for the program manager. It has little, if any, utility to the decision maker. The evaluation may have been built on a model that is fundamentally at odds with the perception of the client, almost guaranteeing it will not be used.

A final important reason for the lack of evaluation use is that government bureaucracies presently provide little incentive for the results to be used. If there is little or no incentive for public officials to evaluate and improve the performance of their programs, then there is little likelihood of enthusiastic application of evaluation results. A program manager who alters his program or suggests its reduction or termination when evaluation indicates program ineffectiveness should be rewarded. Current "reward" systems generally do not operate in this manner (Hatry et al. 1973: 113). As Downs argues, the incentive system in most bureaucracies rewards program continuity and growth (1967).

Although evaluation does not necessarily mean program curtailment or termination, the two are equated in the minds of many. This, of course, returns to the perception that the evaluations must be critical and their results must be damaging to a program. These perceptions, however understandable, need redress. The basic purpose of evaluation is positive (i.e., to improve programs or their services), not negative (i.e., to dismantle them). Once this attitude can be accepted and a set of internal incentives devised to encourage evaluation and program modification based on the results of that evaluation, then this particular obstacle of program evaluation and change will be reduced.

4.10 THE VALUE OF PROGRAM EVALUATION

Program evaluation can be useful in supporting government policy and program planning and implementation, but it is far from being a panacea. In assessing the overall value of program evaluation to government decision making, the benefits, costs, and limitations of the enterprise need to be considered.

4.10.1 The Benefits of Evaluation

Many clear and important benefits can be ascribed to the formal evaluation of government programs, both in terms of specific programs and in a more general policy sense. Feedback from the results of programs can lead to more effective program administration and the improved design of new programs in a similar vein by emphasizing the program outcomes and effects rather than their inputs. Accurate feedback from program impacts enables administrators to adjust program management and implementation strategies to better meet their program's objectives. At the same time, accurate information about program outcomes provides a sounder basis for the annual budget decisions. Under the ever-pressing conditions of fiscal constraint, programs that can document their effectiveness are likely to fare better than those that cannot. Formal evaluation can help identify program externalities, thereby providing recommendations to mitigate the negative impacts of programs and to capitalize on positive side effects.

In a specific programmatic sense, federal, state, and local governments need to share the benefits of their experiences in developing and implementing specific types of solar programs. Without careful measurement and documentation of program performance, most of the value of such experience cannot be transferred. If all federal, state, and local government programs are evaluated and shared formally, the resulting information eventually would indicate which activities within a government program area (e.g., solar energy) are performed most effectively at each level.

In a more general sense, good program evaluation research can contribute to the overall knowledge of political and social behavior. Such knowledge ultimately provides direct benefits to government planning and management and indirect benefits to all the recipients.

4.10.2 The Costs of Evaluation

Evaluation has both direct and indirect costs. The direct costs of an evaluation include the total financial and other material resources, the evaluation staff, and the time required to conduct the evaluation. The resource cost of an evaluation may range from 1% to 15% of total program resources, depending on the nature of the program being evaluated. The number of professional staff required will vary depending on the scope, complexity, and duration of the evaluation task, as will the time required to complete an evaluation.

Evaluation also can entail a variety of indirect costs, such as the opportunity costs of the resources devoted to the evaluation which could be devoted to other government activities. There may be costs of the evaluation's effect on the program in terms of administrative burden, management effectiveness, or even staff morale. The serious costs of the reorganization, reduction, or termination of a program as a result of the evaluation may be incurred. In some cases, the net costs of reorganizing, reducing, or terminating

even a deficient program actually may turn out to be greater than the cost of maintaining the status quo. Finally, there may be a set of political costs. Virtually every program has its sponsors and defenders who will be offended by a critical evaluation; their political ire must be counted, if only vaguely, in the cost of any evaluation.

4.10.3 Limitations of Evaluation

Even under the best circumstances, there are severe limitations that need to be recognized regarding what program evaluation can accomplish. For example, even under good conditions, some questions about program performance cannot be answered at reasonable cost, or even at all. This is especially true when different programs are being compared and evaluated; e.g., how does one evaluate a dollar spent on health insurance versus one spent on a housing or food subsidy (see Rivlin 1971)?

Basically, evaluation is a judgmental art. Some program objectives are so ambiguous or qualitative that they resist any consensual, objective measures. Evaluation data are not equivalent to an objective "truth." Equally skillful evaluators may design evaluations of the same program yielding greatly different results, especially in areas where uncertainties exist. In short, evaluation must operate with limited tools in an imperfect world.

Program evaluation at best indicates what a program is doing, not its social worth. Evaluation is intended to provide some information that may help decide whether a program is worth keeping, changing, or eliminating. Its ultimate worth is a political and social choice upon which formal evaluation has an advisory effect. This again points out that evaluation operates in the political arena and must adapt to the prevailing political environs.

4.10.4 The Future of Program Evaluation

As repeatedly stressed, evaluation is still a young discipline. Although there is much that formal evaluation can accomplish, there is much that evaluation cannot do and perhaps even more that evaluation does not know whether it can do or not. This suggests that evaluation itself is in need of critical evaluation and subsequent refinement. Government officials and professional evaluators both would be served better if they realized that evaluation is a still-developing and imperfect activity with imperfectly known strengths and weaknesses and with results subject to individual, organizational, and social vagaries. Eva Baker noted these conditions as applied to evaluating solar energy programs:

You have to get your constituents to recognize that you see evaluation of solar energy as a developmental process, that you're not coming in with definitive and permanent techniques, that your evaluation plan, technology, and so on, are also developmental and subject to revision, just as are the R&D activities themselves. . . . We're working on these things; we're going to try to provide you as good a feedback as possible; we may be wrong sometimes; sometimes we may hit on the equivalent of a breakthrough. But one of the things I think makes evaluation a lot less threatening is if they see you and you see yourself in that kind of mode.

Some suggestions for evaluation research that could be particularly useful in assessing public policies and programs include research on basic evaluation theory and methodology

(e.g., conceptualization, design, measurement, data collection, and analysis), evaluation case study and documentation, and how to integrate evaluation with other parts of the policy process. These kinds of evaluation development efforts are especially pertinent in the area of solar energy program evaluation because of its unique combination of basic methodology through commercialized technologies, its multiple sponsors and consumers, and its potential sociological and political ramifications. Let us, therefore, turn to a description of the solar energy technology programs so that we can appreciate the special challenges they present to evaluation.



SECTION 5.0

OVERVIEW OF U.S. SOLAR ENERGY PROGRAMS

To place the requirements, limitations, and opportunities of solar energy program evaluation in context, some appreciation of the nature and scope of solar program activities in the United States is necessary. Although a firm figure is difficult to calculate, total federal, state, and local government expenditures on solar energy at present are on the order of \$1 billion/year.

Although solar programs are still a relatively minor portion of all government activity, investment in solar activities has been growing very rapidly; a billion dollars is hardly an insignificant amount of money and that budget is growing. Federal and local government solar programs are numerous, diverse, and expanding. A detailed description of U.S. solar program activity in the public sector far exceeds the scope of this report; this section presents only a brief overview. (For greater detail, see Perelman, ed. 1978.) It does not include a review of the numerous private-sector involvements in the development and commercialization of solar energy technologies.

5.1 WHAT IS SOLAR ENERGY?

The solar energy program covers a wide variety of energy resources and technologies. In addition to the direct absorption or conversion of sunlight to power sources by such methods as photovoltaic cells, solar energy also includes wind energy, biomass (plant or animal products), hydropower (falling water), waves, ocean thermal gradients, ocean salinity gradients, and ocean currents. Of the so-called "renewable" energy sources, only tides and geothermal energy are not generally classified as solar-derived energy sources.

Photovoltaic devices (solar cells) use semiconductor materials to convert sunlight directly into electricity. Solar thermal power systems use the heat of concentrated sunlight to drive turbines for generating electricity. Wind turbines are used to pump water, generate electricity, and sometimes to produce heat. Ocean thermal energy conversion (OTEC) is an experimental technology designed to use temperature differences in some parts of the ocean to generate electricity. Solar powered satellites (SPS) have been proposed to orbit arrays of photovoltaic cells for converting sunlight into microwave beams which could be reconverted to electricity by receiving stations on the earth. Biomass energy is humankind's first and oldest energy source; various plant and animal products can be used to provide gaseous, liquid, or solid fuel, as well as to provide substitutes for energy-intensive materials. Some more exotic technologies now being studied would use chemicals to convert light or heat from the sun into electricity or fuels.

About a dozen general solar energy technologies are now in operation or in some stage of research and development toward practical application. Solar space and water heating is an established industry; solar cooling for buildings still is being developed. Passive solar design of buildings, using windows, building orientation, insulation and various nonmechanical systems to minimize heating and cooling requirements, is being adopted rapidly by architectural and construction firms for both residential and commercial structures. Agricultural and industrial process heat applications provide heat over a range of temperatures for crop drying, fermentation, washing, process steam, and manufacturing processes.

Although the current application of these technologies is low, their potential is significant. The Harvard Energy Policy study projects that solar energy can supply close to 10% of American energy demands by the mid-1980s (Stobaugh and Yergin, eds. 1979). The Domestic Policy Review of Solar Energy estimates that by the year 2000, the United States could use solar energy technologies to meet over 16% of its energy needs, according to the maximum practical market penetration level (U.S. DOE, February 1979: 24), and President Carter recently announced a national goal of 20% by the year 2000.

5.2 TYPES OF SOLAR ENERGY PROGRAMS

A variety of government-sponsored solar energy programs can be arrayed along a spectrum (see Fig. 5-1). Many of these programs are applicable directly to solar energy, as the Domestic Policy Review of Solar Energy makes clear (U.S. DOE, February 1979).

Basic Research explores the basic scientific knowledge required to understand the physical nature of solar energy resources and systems with no direct couple to practical applications.

R&D includes applied technology research and exploratory development. Applied research has some focus on potential applications, though not specifically tied to particular products, while development begins to examine specific technological systems for practical application.

Demonstration programs apply technological systems in a field setting which approximates the conditions under which the technology ultimately is expected to be used. Demonstrations can serve several purposes: to provide an opportunity for further testing and improvement of technical systems under verisimilar field conditions; to explore potential environmental, institutional, and socioeconomic impacts of the deployment of the systems; and to stimulate the interest of entrepreneurs and consumers in the commercial use of systems. In general, then, demonstration programs are intended to reduce the technical performance and economic uncertainties that are inherently a part of new technologies (see Baer et al. 1976 or 1977).

Indicative government measures are not formal programs, but official expressions of goals, or directions which government officials wish to have applied to public or private decisions. Voluntary or semivoluntary guidelines, (e.g., wage and price), general statements of executive policy (e.g., human rights), and legislative resolutions are indicative measures which differ from programs in their absence of formal sanctions, investments, or regulations. An indicative measure in the context of solar energy is the National Energy Plan's call for 2.5 million solar homes by 1985.

Information, Communication, and Education programs serve a wide variety of purposes in the development and diffusion of technology systems, and especially solar energy systems (Roessner et al. 1979 survey this issue). This set of activities includes the preparation and promulgation of university programs for solar scientists and engineers, technician training programs, energy extension service, teacher training, instructional materials development, data bases and information centers, workshops and conferences, speaker bureaus, and printed materials ranging from brochures to books.

Financial incentive programs involve several types of fiscal measures intended to encourage development or adoption of solar energy systems. Such programs may include tax deductions or credits for the purchase of solar equipment; sales or property tax

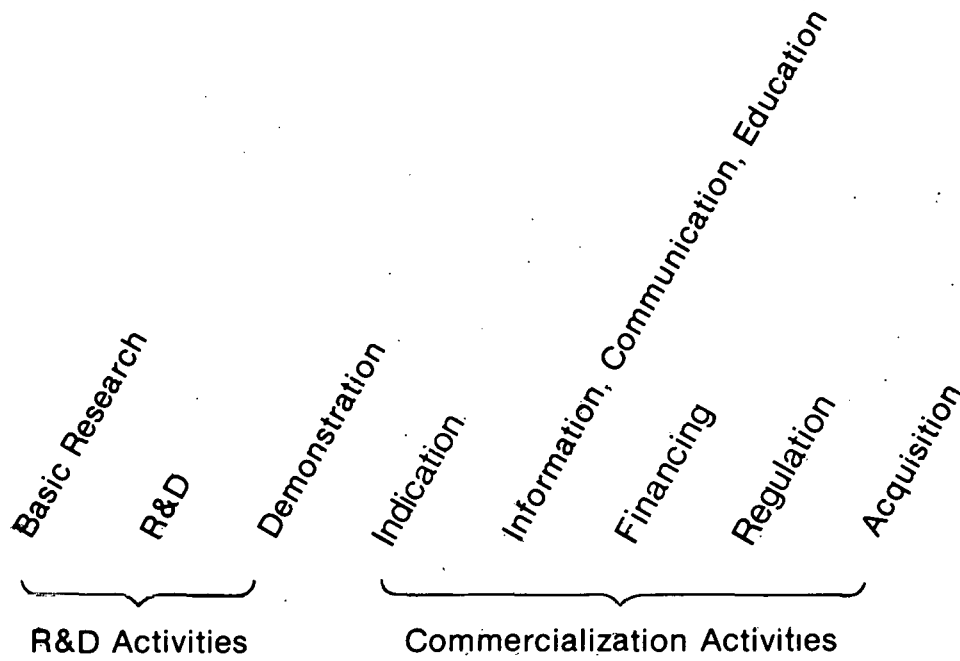


Figure 5-1. Types of Solar Energy Programs

exemptions for the purchase or installation of solar systems; accelerated depreciation allowances for solar equipment; direct grants, rebates, or subsidies; and government insured, guaranteed, or subsidized loans. Financial incentives also may be provided to encourage R&D activities in the private sector.

Regulatory programs can encourage the development and use of solar energy. Federal, state, and local government regulatory practices in transportation, commerce, housing, land use, environmental protection, and health and safety could be adjusted to increase the attractiveness of solar energy relative to conventional energy sources. Conversely, they could act as obstacles to solar energy adoptions.

Acquisition programs involve using the government purchase and operation of new technologies for its own activities as a way to encourage general public acceptance. Acquisition programs would include government procurements of solar energy systems for government buildings or military systems or even giving preference in government contracts to companies using solar energy. The acquisition of solar energy systems by the Department of Defense for supporting its continental facilities has been proposed. Acquisition would also include public works programs for building facilities intended primarily for public use (e.g., airports, water storage, reclamation, or treatment facilities). These programs could enhance the development and diffusion of solar energy systems in three ways: technically, by the construction of publicly-owned solar power generation facilities (e.g., solar thermal electric plants or biogas generation plants), thus improving the technology; economically, because the large government purchase would begin to induce economy of scale reductions in the cost of solar energy equipments; and promotionally, by incorporating solar energy production or use into the design of highly visible public works projects (e.g., solar-heated bus shelters or solar process heat in sewage treatment plants).

5.3 FEDERAL GOVERNMENT SOLAR ENERGY PROGRAMS

The Paley Report (U.S. President's Commission on Materials Policy 1952) and the Cambel Report (U.S. Executive Office of the President 1964) both forecast a major contribution to national energy needs by solar energy. During the 1950s, the National Science Foundation (NSF) sponsored research indirectly related to solar energy and, by 1961, was funding one or two projects directly concerned with solar energy. Total annual solar energy research expenditures through 1970 were only \$100,000. Despite popular and scientific interest in solar energy, no solar energy bills were passed by Congress.

In FY71, the NSF solar energy budget grew to over \$1 million. In 1973, an Atomic Energy Commission panel recommended spending \$200 million over five years for solar energy research. From FY71 to FY73 the solar budget totaled more than \$4 million. During this period, reports by NSF and NASA (1972) and by the Committee on Science and Technology urged the active development of solar technologies as future energy sources. Following the Arab oil embargo in 1973-1974, the Project Independence Blueprint for Solar Energy (November 1974) and the NSF National Solar Energy Program (December 1974) were published. Congress enacted several solar energy laws, including the Solar Heating and Cooling Demonstration Act and the Solar Energy Research, Development, and Demonstration Act.

Also in 1974, Congress dismembered the Atomic Energy Commission and distributed its energy responsibilities to the newly created Federal Energy Administration (FEA) and the U.S. Energy Research and Development Administration (ERDA). ERDA's programs for

solar energy RD&D were described in A National Plan for Solar Heating and Cooling (ERDA-23), A National Plan for Energy Research, Development, and Demonstration: Creating Energy Choices for the Future (ERDA-48), and Definition Report: National Solar Energy Research, Development and Demonstration Program (ERDA-49). ERDA solar energy projections were generally cautious, estimating that solar energy could supply only 7% of national energy demands by year 2000 and 25% by 2020, assuming significant cost reductions. ERDA's planning emphasized solar heating and cooling in the near term and solar electricity technologies as a high priority for the longer term.

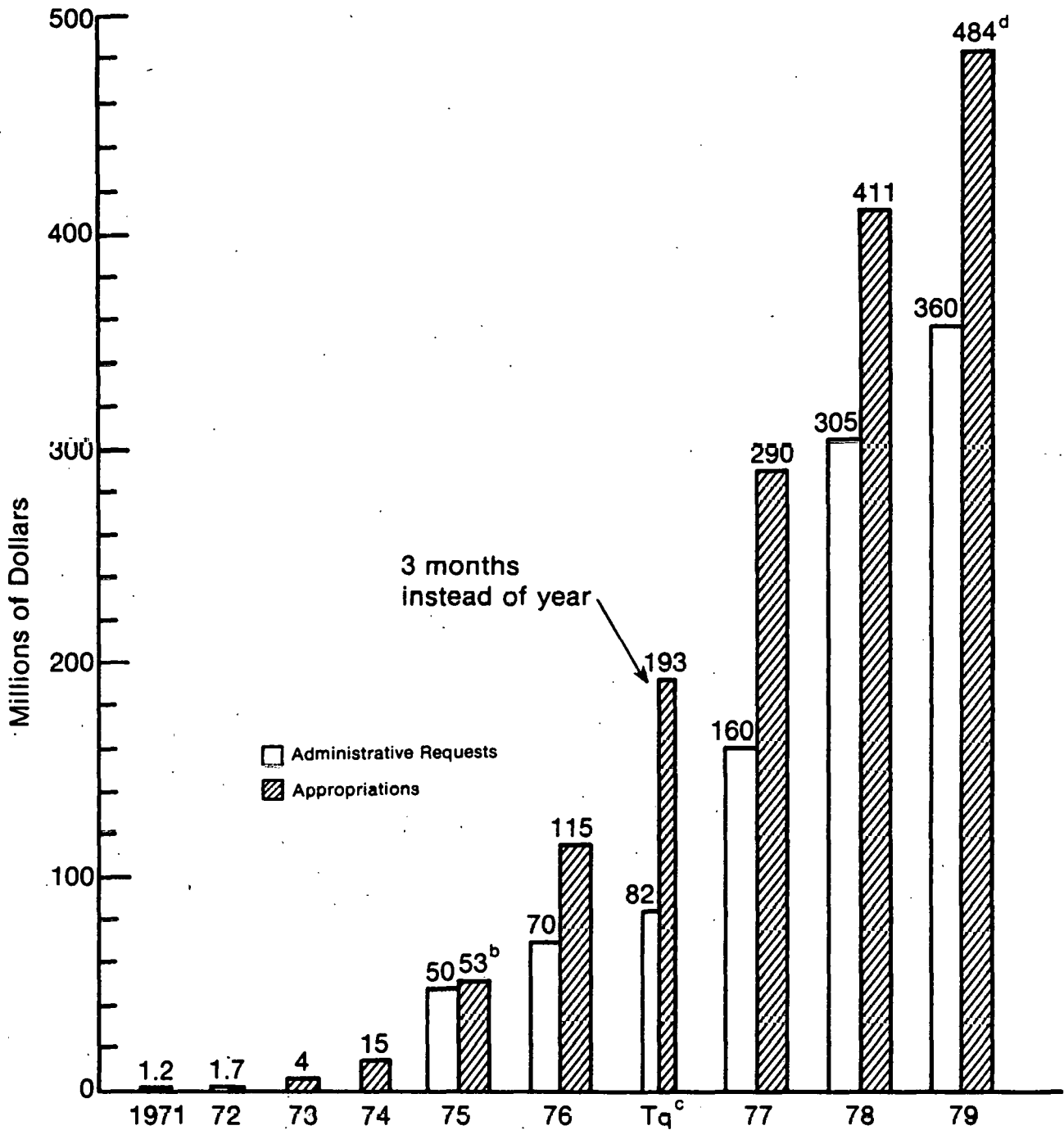
President Carter's National Energy Plan (NEP) of April 1977 established a national goal for solar energy of 2.5 million solar-heated homes by 1985. The general policies articulated by the NEP encouraged the development and use of solar energy. Also in 1977, at the President's initiative, Congress consolidated FEA, ERDA, and several other federal energy agencies into the U.S. Department of Energy (DOE). Acting under the congressional mandate of P.L. 93-473, DOE established the Solar Energy Research Institute (SERI) and four Regional Solar Energy Centers (RSEC) to promote solar energy development and commercialization. In 1978, Congress passed the National Energy Act, which included provisions for a tax credit on the purchase of certain types of solar equipment. During this time, both administration budget requests and Congressional allocations for solar energy programs grew dramatically. In fact, as Fig. 5-2 displays, congressional approval of funds exceeds administration requests.

5.3.1 Department of Energy (DOE)

Until 1977, the majority of federal efforts in solar energy R&D and commercialization was carried out by ERDA, FEA, and the U.S. Department of Housing and Urban Development (HUD). DOE has consolidated the responsibility for the following solar energy activities:

- develop and implement general policies and specific programs to encourage the widespread use of solar energy;
- consult with representatives of science, industry, environmental organizations, and consumers interested in solar energy;
- design mechanisms for the effective participation of state and local governments in the resolution of solar energy problems;
- research, development, and demonstration of solar energy technologies;
- determine the solar energy resource base; and
- coordinate solar energy information dissemination efforts.

Fig. 5-3 shows the organizational location of DOE's solar-related activities. Although the DOE Offices of the Assistant Secretaries for Conservation and Solar Applications (CSA) and Energy Technology (ET) are the focal points of solar efforts, many other offices and divisions contribute to the DOE solar program. In FY77, the Materials Science Program (now in the Office of the Director for Energy Research) funded about \$2 million in research areas related to solar energy. The Office of the Assistant Secretary for Environment (EV) conducts solar energy environmental R&D, reviews DOE policies and strategies for environmental impacts, and prepares environmental impact statements. The Energy Information Administration (EIA) collects, analyzes, and disseminates solar energy information. Solar energy policies are evaluated by the Office of the Assistant Secretary for Policy and Evaluation (PE).



^a Administration requests available only for FY75-FY79; appropriations not yet made for FY79.

^b Also reported as 48.9.

^c Transition quarter between FY76 and FY77; bars show what funding would have been if it had been for a full fiscal year.

^d Does not include \$28.0 million appropriated for small-scale hydroelectric.

Source: Data from DOE; Congress, House of Representatives: Committee on Nuclear and Alternative Energy Systems.

Figure 5-2. Solar Budget: Total Administration Requests and Appropriations, 1971-80^a

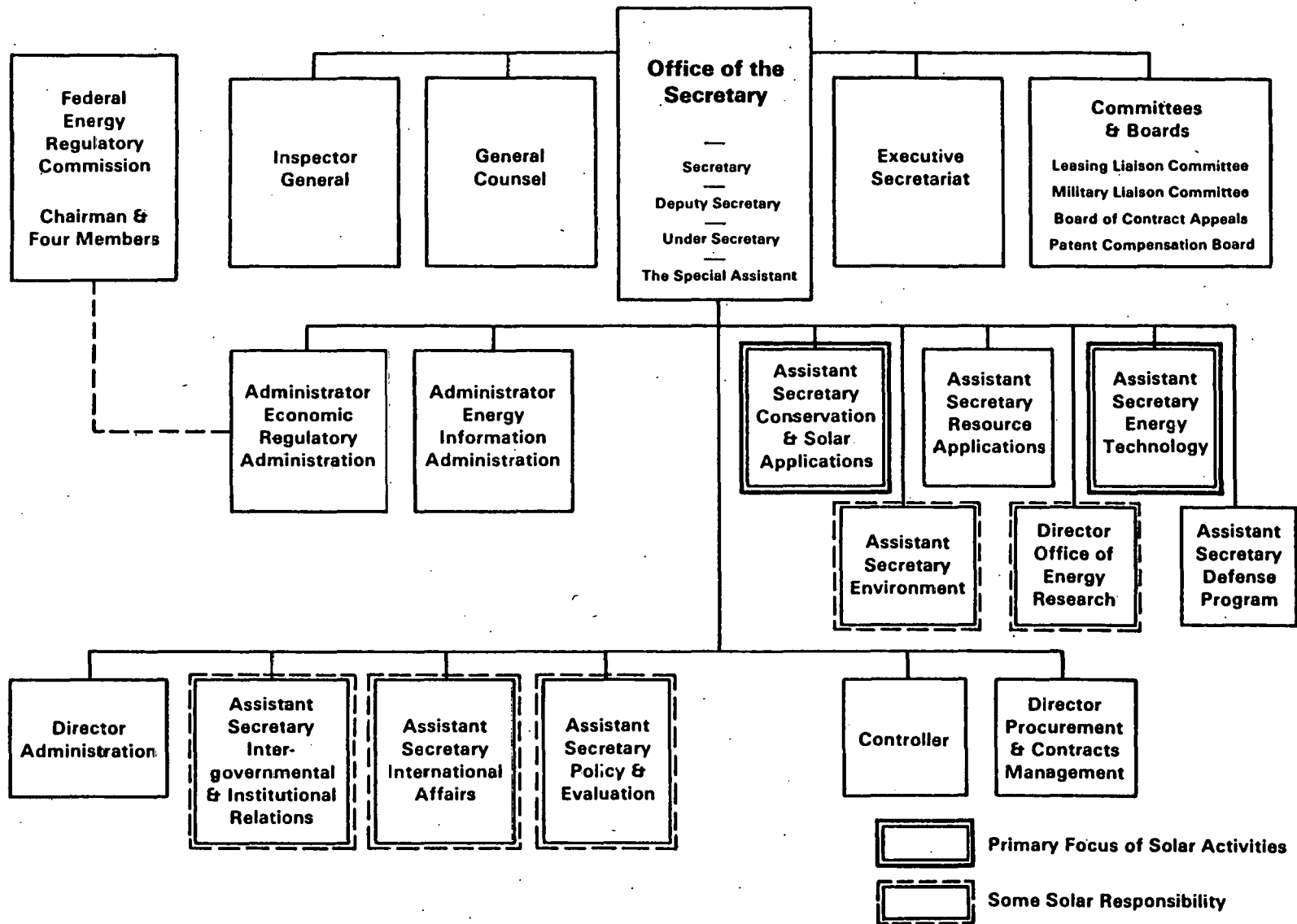


Figure 5-3. Department of Energy

Most of DOE's solar energy programs are divided between CSA and ET on the basis of technology development status. Principal responsibility for solar technologies which are considered to be still in the technology development stage (e.g., solar thermal, photovoltaics, wind energy, biomass, and ocean thermal) is delegated to ET. Responsibility for currently demonstrable technologies (solar heating and cooling, agricultural and industrial process heat) is the institutional domain of CSA. As a particular solar energy technology progresses from R&D to commercialization, program lead responsibility shifts from ET to CSA. The organizational structure of DOE solar energy programs, then, focuses primarily on individual technologies rather than on functional categories such as energy end-use demands or strategic requirements.

SERI is administered by ET while the Regional Solar Energy Centers are administered by CSA. The national laboratories (e.g., Argonne, Brookhaven, Oak Ridge), which carry out a major portion of the DOE solar energy R&D effort, report to the Office of Energy Research. The national laboratories performed an estimated \$51.3 million of RD&D in support of DOE solar energy programs in FY78. Various NASA field centers are involved in specific RD&D activities supporting DOE solar heating and cooling, wind energy, photovoltaics, and thermal power systems programs. The budgets of these institutions are not line items in the overall DOE budget but are charged to the program budgets.

5.3.2 Other Executive Departments

HUD manages jointly with DOE the residential active solar heating and cooling demonstration program, which includes the development and demonstrations of solar technology, the definition of standards and certification procedures for solar energy systems, and the dissemination of information on solar energy heating and cooling systems. In 1976, the HUD Community Services Administration established the National Center for Appropriate Technology in Butte, Mont., to provide technical assistance and small grants for low-income, appropriate technology projects.

The U.S. Department of Commerce is responsible for several solar energy programs, such as providing technical assistance and funds for planning, research, demonstration projects, and training programs which include some solar energy activities. The Center for Building Technology of the National Bureau of Standards is developing a set of performance criteria for solar energy equipment and residential dwellings, monitoring the design, testing, evaluation, construction, and operational phases of the DOE/HUD demonstration programs, and setting interim standards for solar heating systems. The National Oceanic and Atmospheric Administration has been tasked to develop and conduct a plan for the inventory of all forms of solar energy resources associated with federal lands.

The U.S. Department of Defense (DOD) is coordinating the installation of solar heating and cooling systems in DOD facilities. DOD also is providing 10 military bases for solar electric applications where they satisfy military needs and are competitive with conventional systems under the terms of the Military Construction Act of 1978. The U.S. Department of Agriculture (DOA) manages agricultural aspects of the DOE agricultural and industrial process heat program, as well as conducting R&D in agricultural applications of solar energy. Other DOA solar energy programs under consideration include R&D grants, a research information system, a model farm program, a demonstration program, and several regional R&D centers.

The U.S. Department of State (DOS) has responsibility for negotiating international agreements on bilateral or multilateral RD&D projects and on potential large-scale

export of commercial solar energy projects. The International Development and Food Assistance Act of 1977 broadened the DOS role in solar energy by authorizing cooperative energy programs with developing countries to emphasize appropriate, small-scale, decentralized, renewable energy sources for rural areas. The Agency for International Development (AID) is undertaking a major effort in solar energy projects in its foreign assistance programs; projects totalling in excess of \$80 million have been identified or funded (Ashworth 1979).

5.3.3 International Programs

The federal government cooperates in a number of solar energy programs with international agencies and foreign governments. The United States is one of 19 member nations of the International Energy Agency (IEA), which was created in the wake of the Arab oil embargo of 1973-1974. Cooperative R&D projects are underway in solar heating and cooling and solar thermal technologies. IEA nations are considering joint projects in wind energy conversion, wave power, and biomass systems. The 19 member nations, including the United States, also participate in the Committee on the Challenges of Modern Society (CCMS). A Solar Energy Pilot Study, initiated by CCMS in 1973, has led to the development of an international data format for reporting results and progress of programs in solar heating and cooling of buildings, and a systems performance information exchange network.

The United States has bilateral agreements for cooperative projects in solar energy with France, the USSR, Spain, Japan, Denmark, and Saudi Arabia involving a variety of RD&D projects in high-temperature process heat, solar thermal systems, solar heating, photovoltaics, collectors and component materials, wind energy conversion, and other technologies (see Perelman, ed. 1978: Chap. 5).

The United States is also taking a leading role in the renewable energy programs within United Nation agencies. In part, this is in preparation for the U.N. Conference on New and Renewable Sources of Energy, to be held in New York in 1981. The United States also is a major contributor to the World Bank which is encouraging lesser developed nations to develop their solar energy resources.

In addition to intergovernmental projects, some institutional international programs primarily with universities are supported by some federal funding. These activities are primarily basic research projects in photovoltaics, solar heating and cooling, solar thermal power, and flat-plate and concentrating collectors. Some educational projects also exist. Sponsoring and/or supporting institutions are NSF, DOE, the U.S.-Israel Binational Science Foundation, and the International Energy Agency.

5.4 STATE GOVERNMENT SOLAR ENERGY PROGRAMS

State governments have initiated a wide variety of programs to promote the development and use of solar energy. In addition to legislative and regulatory actions, state organizations have been created for a wide variety of purposes, such as to collect and disseminate solar energy information, to revise building codes, to organize conferences, and to design solar energy curricula. Both the number and substance of these initiatives vary from state to state but generally include financial, regulatory, RD&D, and information programs. Table 5-1 summarizes the status of state solar energy financial incentive and regulatory programs through 1977. State programs are discussed at length in Ashworth et al. (1979).

Table 5-1. SOLAR RECORD, STATE BY STATE

	Income Tax Incentives	Property Tax Incentives	Sales Tax Incentives	Grants or Loans	Provision Allowing Solar Easements	Equipment Standards Construction Rules
Alabama						
Alaska	X					
Arizona	X	X	X			
Arkansas	X					
California	X			X	X	X
Colorado	X	X			X	
Connecticut			X			X
Delaware	X					
Florida			X		X	X
Georgia			X		X	
Hawaii	X	X				
Idaho	X				X	
Illinois		X				X
Indiana		X				
Iowa		X		X		
Kansas	X	X			X	
Kentucky						
Louisiana		X				X
Maine		X	X	X		
Maryland		X			X	
Massachusetts		X	X			
Michigan	X	X	X			X
Minnesota	X	X			X	X
Mississippi			X			
Missouri						
Montana	X	X				
Nebraska						
Nevada		X				X
New Hampshire						
New Jersey		X	X		X	X

Table 5-1. SOLAR RECORD, STATE BY STATE (concluded)

	Income Tax Incentives	Property Tax Incentives	Sales Tax Incentives	Grants or Loans	Provision Allowing Solar Easements	Equipment Standards Construction Rules
New Mexico	X				X	X
New York		Y				X
North Carolina	X	X				
North Dakota	X	X			X	
Ohio						
Oklahoma	X					
Oregon	X	X		X		X
Pennsylvania						
Rhode Island		X				
South Carolina						
South Dakota		X				
Tennessee		X		X		
Texas		X	X			
Utah						
Vermont	X					
Virginia					X	
Washington		X				
West Virginia						
Wisconsin	X					
Wyoming						

^a Local option authorized by state.

Source: National Solar Heating and Cooling Information Center in New York Times (13 August 1979: D-8).

5.4.1 Financial Incentives

State financial incentives now in existence include partial or complete exemptions from increased property taxes resulting from solar installations, income tax credits or deductions, exemptions from sales and use taxes, exemptions from franchise taxes, and low-interest loans. The design, magnitude, eligibility, and administration of such programs vary greatly from state to state. Colorado allows a 100% deduction of the cost of solar equipment from the state taxable income. California provides a 55% state income tax credit which cannot be added to the federal tax credit while Arizona offers a 35% tax credit which can be added to the federal tax credit. Texas provides exemption from sales tax on receipts from the sale, lease, and rental of solar devices (Ashworth et al. 1979: Chap. 1).

5.4.2 Regulatory Measures

State regulatory programs designed to promote the use of solar energy include standards, energy consumption analysis, energy use disclosures, and solar rights. Again, these measures vary greatly among states.

Standards legislation addresses issues in building design, building products, and solar equipment. Legislation mandating the development and enforcement standards has been enacted. Some states, including California and Florida, have adopted performance-oriented approaches, but most use prescriptive or component performance criteria (Odland in Ashworth et al. 1979).

Legislation requiring energy consumption analysis or energy feasibility studies has been enacted in a number of states. For example, Oregon requires that solar energy alternatives should be considered in comprehensive planning. In Connecticut, life cycle cost analysis of alternative energy systems is a prerequisite of any major state capital project. Utah and Oregon have enacted legislation requiring full disclosure of data on energy consumption and cost to potential purchasers of buildings. Regulation and protection of access to sunlight is an important issue given the fixed, long-term, and capital-intensive nature of most solar installations. Procedures for creating voluntary solar easements or for including solar energy considerations in zoning and land use planning have been enacted in a number of states including California, Colorado, Maryland, New Mexico, and Oregon (Pollock in Ashworth et al. 1979).

5.4.3 RD&D and Information Programs

State RD&D programs generally focus on solar energy technologies which are suited particularly to state requirements or resource opportunities. For instance, states with abundant sunshine are interested in solar process heat applications while those with extensive forests, marshlands, or agricultural residues are interested in developing biomass energy. The RD&D programs include contracts with large research firms or universities as well as small grant programs for individuals or grass roots organizations. The activities supported range from basic laboratory research to systems design to large-scale demonstration projects (Green in Ashworth et al. 1979).

State information and education programs dealing with solar energy are too diverse and extensive to catalogue easily. Many of these solar energy activities are combined with

energy conservation and general energy information dissemination programs, and complement federal government programs (Saltonstall in Ashworth et al. 1979).

5.4.4 State Program Organization

Several states have established solar energy offices with explicit responsibility to administer and/or coordinate state solar programs. Some state solar energy offices are independent entities while others are part of a larger, pre-existing division. For example, Arizona created a Solar Energy Research Commission (ASERC) in 1974 to conduct RD&D program. ASERC not only administers RD&D activities but has taken an important coordinating role in other state solar energy programs (e.g., the tax credit program administered by the Department of Revenue). Minnesota's Solar Energy Office is part of the Minnesota Energy Agency, and advises and coordinates other state solar programs while administering one of the R&D programs. California's state Energy Commission includes divisions with express responsibility for solar energy; a solar energy office also exists within the Office of the Governor.

In most states, solar energy programs are implemented and administered by several state agencies, and no formal state solar office may exist at all. Legislation establishing state solar energy programs sometimes delegates responsibility jointly to two or more agencies, or neglects to define administrative responsibility adequately.

State solar energy associations and other nongovernmental organizations often play an important role in the creation, design, and even implementation of state solar programs. Such organizations lobby for and may help write solar energy legislation, assist in the design of regulations or of eligibility requirements for R&D grants and contracts, provide support to information or education activities, and act as citizen evaluation groups for solar energy programs.

5.5 LOCAL GOVERNMENT SOLAR ENERGY PROGRAMS

This report will not explicitly address the evaluation of local government solar programs because an adequate treatment of the unique problems of local government program evaluation is beyond the scope of the present project (see Clarren et al. 1978; Perelman in Ashworth et al. 1979). Still, it should be recognized that federal programs can take advantage of the state and local evaluations of their solar energy programs as sources of new and creative ideas, operations, and evaluation techniques (see U.S. House of Representatives, Subcommittee on Energy and Power 1978).

The nation's cities, counties, regional agencies, and local jurisdictions have some important opportunities to affect the development and use of solar energy. Local governments have major responsibilities in land use planning. A number of states have passed enabling legislation giving municipal and county governments authority to include solar energy considerations in comprehensive planning. For example, the Arizona legislature has given broad authority to local authorities to regulate access to sunlight. Los Alamos County, New Mex., provides protection of solar access rights. Cerritos, Calif., adopted an ordinance that incorporates provisions for the installation of solar equipment into the local building codes.

The powers of local governments are limited. Federal and state governments retain direct authority affecting most regional and local solar energy activities. Local governments generally lack resources to fund and support large-scale programs. Nevertheless,

local governments can accomplish much. For instance, grass roots efforts in Davis, Calif., led to the incorporation of solar technologies in the building code over the objections of the building industry; recent evaluations have been uniformly favorable, including positive comments from the construction industry. San Luis, Colo., which has probably the highest per capita use of solar heating in the United States, is a dramatic example of the great potential of community-based solar development. Ultimately, then, local governments can have an important influence on the form and scope of solar energy development.

SECTION 6.0

THE PRESENT CONDITION OF SOLAR ENERGY PROGRAM EVALUATION

Evaluation of solar energy programs at both the federal and state levels is still in its incubation, largely because most of the programs are very new. Also, agencies and officials responsible for most solar programs commonly have little experience in formal evaluation. This chapter gives an overview of the present status of solar energy program evaluation at the federal and state levels.

6.1 BACKGROUND OF FEDERAL PROGRAM EVALUATION

The roots of formal evaluation of federal programs probably lie with the public health programs in the 1920s (Suchman 1967). It became a dominant concern in the federal bureaucracy during the early 1960s, largely due to the efforts of Robert McNamara in the Department of Defense (Enthoven and Smith 1971). McNamara, representing a new generation of systems-oriented management, instituted "cost-effectiveness" criteria in DOD decision making, created an Office for Systems Analysis, and installed a planning, programming, and budgeting system (PPBS), which became a key to government budget preparations (Hitch and McKean 1960; Novick 1975). Inspired by McNamara's management example at DOD, President Johnson, in 1965, required all federal agencies to implement PPBS in the hope of improving the quality of decision making in domestic programs (Rivlin 1971). Most of these other applications of PPBS were less successful than at DOD (Wildavsky 1974; Drew 1967). One exception was the newly created Office of Economic Opportunity (OEO). Rather than becoming extinct, these approaches evolved into new government management systems such as management-by-objectives (MBO) and zero-base budgeting (ZBB).

OEO personnel promulgated the PPBS approach and tried to apply cost-effectiveness and efficiency criteria to social program planning. The OEO staff soon discovered, though, that there were significant differences between military planning and a "war on poverty" (Levine 1970). DOD systems analysis methods were almost exclusively concerned with future requirements and capabilities which could be translated, however arbitrarily, into hardware or technical options. Domestic programs, however, were based largely on the theories of social science and social problems (e.g., poverty, crime, disease) which were less amenable to systems analysis and cost-effectiveness criteria. Tradeoffs between programs were particularly difficult (Rivlin 1971). As a result, OEO's concept of program cost effectiveness eventually evolved to place a major emphasis on evaluations of actual program outcomes.

OEO evaluation was initially based on management audit information, which soon proved to be an unreliable measure of program performance. In 1968, an Evaluation Division was created in the OEO Office of Research, Plans, Programs, and Evaluation (RPP&E), after 1969, called the Office of Planning, Research, and Evaluation (PRE), whose purpose was to evaluate important OEO programs. The primary concern of RPP&E was to provide information supporting budget allocation decisions; responding to the demands of various parts of OEO, RPP&E created several types of evaluation. Following the OEO example, formal program evaluation divisions were established in other federal agencies concerned with domestic social programs.

Another important influence on the evolution of federal program evaluation was the Elementary and Secondary Education Act (ESEA) of 1965. Senator Robert Kennedy

attached an amendment to ESEA which required that one percent of the Title I funds granted to local school agencies be used for evaluation of funded programs, to insure that ESEA funds resulted in measureable benefits for students. He was supported in this amendment by the Office of Education which had just established an evaluation unit (McLaughlin 1975; Meranto 1965). The ESEA evaluation efforts were largely unsuccessful. Thousands of projects were required to be evaluated by grant recipients, who typically had little knowledge of evaluation theory or methodology. Consequently, the evaluations had questionable validity. Since Title I did not specify how the evaluation results were to be used or by whom, the evaluations were largely ignored. Nevertheless, the ESEA experience did not prevent evaluation from becoming a standard procedure in many federal education and other social programs.

In 1968, HUD asked the Urban Institute for an assessment of federal evaluation policy. In their report, Wholey and his colleagues presented a discouraging picture of federal evaluation in the late 1960s.

There is nothing akin to a comprehensive federal evaluation system. Even within the agencies, orderly and integrated evaluation operations have not been established. Funding has been low. Staffing has been worse, forcing undue reliance on outside contractors by agencies that lack the in-house capacity to monitor contract work. The most clear-cut evidence of the primitive state of federal self-evaluation lies in the widespread failure of agencies even to spell out program objectives. Unless goals are precisely stated, there is no standard against which to measure whether the direction of a program or its rate of progress is satisfactory.

The impact of activities that cost the public millions, sometimes billions of dollars has not been measured. One cannot point with confidence to the difference, if any, that most social programs cause in the lives of Americans. It has not been established that one approach has been more effective than another in reducing poverty, eliminating slums or providing quality education to all children. Why the same type of project seems to succeed in one community but fails in another has not been determined. Lack of a solid scientific information base about past and present programs poses severe limitations on the federal government's ability to map out sound future programs. (Wholey et al. 1976: 15-16)

Program evaluation is now a major item of federal spending. A survey by the U.S. Office of Management and Budget (1977) showed that in FY77 over \$243 million were obligated for program evaluation. Of this amount, about 70% was for contracts or grants to conduct program evaluations. Federal agency staffs devoted to program evaluation represented 2,167 staff years (inclusion of GAO staff would increase this total by about 80%), or about 0.11% of total agency personnel; the average cost (salary and benefits) of program evaluation staffs was \$28,100. The bulk of this effort was in social program areas (e.g., in HEW). In the energy area, DOE's predecessor agencies (such as ERDA and FEA) devoted 127 staff years and \$6.98 million in grants and contracts to program evaluation. As an indication of the scale of program evaluation activity in the past decade, a computer search of the Educational Resources Information Center (ERIC) data base revealed over 10,000 citations on the subject of evaluation.

Despite extensive growth in resources and technical improvement in federal evaluation practice during the past ten years, the present condition of federal evaluation probably is not drastically different from Wholey's assessment. Most federal departments now have

some formal evaluation procedures, but these vary greatly from department to department and even from agency to agency according to the particular types of programs and information requirements that exist, although such diversity seems appropriate. Evaluation staffs have grown, but much evaluation work still is performed by contractors. With a few exceptions, the effects of most government activities and their underlying causes generally are not well understood and often cannot be determined.

The effect of evaluation itself is uncertain. Large-scale evaluation studies sometimes have stimulated intense controversy, but their direct impact on policy making is less obvious. An often-cited example is a technically rigorous evaluation of the Head Start program produced by Westinghouse (1969), which showed that the Head Start program had been ineffective in achieving its goal of improving the IQs of minority students. The report stimulated considerable controversy and even was mentioned in a Presidential address. Supporters of the Head Start program argued that the evaluation simply had failed to measure many of the program's social and psychological benefits other than IQ enhancement. In either case, the Head Start program continues.

In 1976, a symposium on "The Use of Evaluation by Federal Agencies," sponsored by the National Institute for Law Enforcement and Criminal Justice, assembled evaluation experts and federal officials to review the state-of-the-evaluation-art. The ensuing discussions (reported in Chelimsky 1977) revealed four major problems in the existing practice of federal program evaluation. First, the relevance of the social research model to agency requirements in the evaluation product was questioned. Second, the perceived failure of present bureaucratic incentive structures to reward agency managers for the effectiveness of their programs or for their efforts to improve that effectiveness via evaluation was seen to impede formal and rigorous evaluation. Third, the difficulty of reconciling differing evaluation perspectives, expectations, and information needs among different evaluation users introduced considerable uncertainty as to the validity of the results. And fourth, program managers found it awkward to select evaluation approaches which simultaneously minimized threats to the program and maximizing the amount of useful information.

Despite a history of uncertain success and some disappointment, federal program evaluation seems to face increasing interest and demand. For example, the U.S. Senate passed the Federal Program Review Act of 1977 which would have established a federal "sunset" regulation. The bill called for a 10-year schedule for the reauthorization of all federal programs, with certain specific exemptions, and specified that no funds could be spent for any program unless the expenditure has been reauthorized by Congress. The bill required authorization bills to set forth sufficient information to permit the comprehensive evaluation of programs. It also required the President to recommend programs to Congress for indepth evaluation along with his own evaluation of the program selected by Congress for evaluation (U.S. Senate 1978). Similar legislation is expected to be introduced in future sessions of Congress, although there are important reservations as to the ability to carry out the necessary evaluations (Brewer 1978; Rivlin in U.S. Senate Subcommittee on Intergovernmental Affairs 1977).

6.2 CHARACTERIZING THE NATIONAL SOLAR ENERGY PROGRAM

Given the mixed record of evaluation in the federal government next to the growing demand, it would be useful to review the prevalent RD&D models against which the solar energy programs are likely to be measured. As emphasized earlier, the appropriate approach to program evaluation is highly dependent on the levels and types of programs

to be evaluated. The collective national solar energy effort is comprised of a variety of different types of programs carried out by diverse government agencies. Still, a general model of solar energy technology RD&D programs is useful to identify broad-gauge evaluation requirements.

At the federal level, the explicit or tacit model of the solar energy program that seems to prevail is of an RD&D program primarily concerned with the research and development of an innovative technology whose measures of success are mainly technical and economic performance. In general, the model is patterned after similar activities in the military, space, and nuclear power programs. Until FY80, with the advent of solar tax credit, the major share of the total government resources invested in solar energy programs was devoted to basic R&D activities. Only in the last few years has there been significant funding in the general commercialization segment of the program. A second reason for this orientation is that many federal officials involved in the solar program were previously engaged in military, space, or nuclear power R&D activities and came to the solar program from DOD, NASA, or the AEC/ERDA. Psychology and organization theories suggest that one might expect officials with extensive backgrounds in military, space, or nuclear power programs to treat their new responsibilities as extensions of their prior experience. Similarly, the perception of solar program goals largely as a function of hardware development towards specified performance characteristics and threshold economic costs is highly consistent with program goals in the military, space, and nuclear power R&D programs.

One needs to ask if this model is consistent or divergent from the emerging and future direction of the national solar energy effort. If it is inconsistent, then evaluation models for earlier RD&D programs may prove inadequate for assessing the various solar energy programs. Upon reflection, three facets of the posited RD&D model fail to correspond to the existing and developing nature of the overall national solar energy RD&D programs. First, unlike the space and military RD&D programs, the national solar program is no longer exclusively a federal RD&D program, nor even as a federal activity is it predominantly in the domain of a single federal agency or department, as was the case with the nuclear reactor. Decentralization has become a key feature of the solar program, not only in terms of its constituency's interest in the scale, distribution, and use of the technologies, but also in terms of the governance of and the participation in the program. Several federal agencies, state and local governments, nongovernmental organizations, and international agencies have developed a substantial and increasing role in the initiation, design, and conduct of U.S. solar energy programs. The decentralization and diffusion trends of solar energy program activities can be expected to intensify. Partly this is because solar energy is amenable to relatively low-cost, decentralized efforts, partly because solar energy resources are diverse and widely distributed, and partly because solar energy is a social innovation which potentially affects the interests and demands of a broad variety of stakeholders (see Clarren et al. 1978).

Second, the image of the solar energy program as primarily a technology R&D program is now obsolete. The R&D proportion of the total program effort, while still important for some solar technologies, is generally declining for many solar applications. Concerns for the commercialization and diffusion of appropriate solar technologies which are considered ready or nearly ready for widespread practical utilizations are increasing (Commoner 1979). The bulk of solar energy legislation pending at both federal and state levels would substantially increase investment and activity in the commercialization and diffusion types of programs. For example, legislation passed in the last session of Congress calls for the investment of almost \$100 million for federal procurements of photovoltaic systems; similar funds were authorized for solar heating systems. Both programs are directed towards the commercialization of these systems. Congress has enacted an income

tax credit for solar heating systems as part of the National Energy Act. Although the actual cost of this program to the federal budget is difficult to estimate, the scale of the program certainly constitutes a major departure from previous government technology RD&D initiatives, especially the military or space programs which were not constrained by market criteria.

Third, the DOD and NASA experiences might be considered as inappropriate models for the national solar energy development program. In both cases, the development of advanced technology has been directed at end uses distinct from commercial or civilian activities. The clients for the products of military and space R&D have been the same agencies that were conducting the R&D effort. The functions of the technologies generally have been well defined. Commercial economic factors have a minimal role in the definition of program objectives. Both DOD and NASA programs have sought to transfer technologies developed for military or space application to commercial markets (see Bean and Roessner 1978; U.S. Comptroller General 1972). Still, the primary goal of the military and space R&D programs has been to meet the operational requirements of the sponsoring agencies.

The contrast of the solar energy programs with the previous DOD and NASA programs should be emphasized. Solar energy programs include social, environmental, and economic components which go well beyond the technical performance and operational cost characteristics of the technology's structural elements. In addition, market tests must be passed. Both of these characteristics distinguish solar energy RD&D programs from their DOD and NASA counterparts.

The nuclear power RD&D program is a somewhat more relevant model for the solar energy program, but still is not completely satisfactory (Hammond and Metz 1977). Like solar energy technologies, the nuclear reactor RD&D programs had the goal of the development and widespread application of an innovative energy technology. The nuclear power reactor RD&D program did emphasize the critical importance of anticipating, assessing, and accommodating the social, psychological, political, economic, and ecological components of the desired technological innovation (see Perry et al. 1977). Still, there are several reasons why the nuclear power program is an inappropriate model for the solar energy RD&D programs. The nuclear program has focused for three decades on essentially a single technology; the solar program includes at least a dozen very different technologies. Nuclear power is designed to serve basically only a single end-use application—the generation of electricity. Solar technologies are at least potentially capable of serving a wide variety of energy end-use requirements. Finally, the economics of nuclear power depend on very large-scale, centralized facilities, which contrast sharply with the decentralized character of the solar energy program.

Solar and other new energy programs are relatively novel to state government. The states' prior principal experience in energy matters has been in the regulation of public utilities and in the taxation and regulation of energy resource development. To a significant extent, the states' recent and growing involvement in solar energy R&D, commercialization, and diffusion is unprecedented. If there is any image of the state solar energy program, it is an eclectic model built from a variety of state government interests: consumer protection, environmental protection, desire to capitalize on indigenous state energy resources and reduce dependence on outside sources, interest in internal economic development, and the wish to reduce the impacts of growing exploitation of a state's fossil energy resources.

In summary, an appropriate characterization of the national solar energy program must recognize that the program combines elements of a particular type of R&D

activity—what Rubenstein (1977) has labeled "federal civilian-oriented research and development" (FCR&D)—with elements of a program for intentional diffusion of innovation (Roessner et al. 1979). This combination calls for a new and particular blend of evaluation skills.

If the applicability of the dominant RD&D models from DOD & NASA is questionable, other technology RD&D research might be more relevant. In their research for the Experimental Technologies Incentives Program (ETIP), Norm McEachron and his colleagues (1978) reviewed a large number of FCR&D programs, such as a clean boiler fuel demonstration project for enhanced oil well recovery and copper extraction by a lime roast leach method. Walter Baer and his colleagues at The Rand Corporation also studied a number of ETIP programs to determine what were the appropriate objectives for federal demonstration programs and the policy means to those objectives (Baer et al. 1976). These evaluations of relatively small-scale, civilian technology development programs may provide some models for similar solar energy activities.

Government programs for the widespread diffusion of civilian technologies have limited precedents in the United States. Given the political tradition of "laissez faire," American government has sought to facilitate entrepreneurship rather than to promote adoption of technological innovation. This is not to say that U.S. government agencies have not played a pivotal role in diffusion of technological change; government regulations, patents, subsidies, and investments have been critical to the development and diffusion of railroads, automobiles, electricity, radio and television, and aviation. But the government role in these technologies more often has been reactive and supportive than intentionally developmental. A significant exception is in public health.

Perhaps the most important historical model of government diffusion program of a civilian technology is the national agricultural extension program. For over a century the federal agricultural program has developed and disseminated decentralized technology innovations, tasks the solar technology development program face. Both programs share several common issues, such as the appropriateness of the technologies, the trade-off between large- and small-scale applications, the degree of centralization of both program and institutional structures, the interaction and roles of public agencies and private industry, and the need for an integrated process of basic research, technology development, commercialization, and diffusion. Still, the extent of the agricultural program analogy to the solar program is limited. For most of its history, the agricultural program has been directed at a well-defined, largely homogeneous, rural clientele; the solar program's intended clients are far more demographically varied. The agricultural program has been production oriented whereas the solar program needs to be responsive to several types of energy end-use consumption. The solar program also is broader in scope, including components similar to social welfare, community development, and environmental protection programs.

Despite several reservations, the FCR&D and agricultural program models share sufficient characteristics with the solar energy program that they can suggest some concepts of success as a prelude to program evaluation criteria. McEachron et al. (1978) evaluated the processes of planning and project initiation, project selection, project deployment planning, project implementation and administration, and project review and evaluation in terms of producing a new commercial technology. Of these processes, they found that planning and project initiation was most important to eventual program success. Within the area of FCR&D planning and project initiation, they then identified several factors which they claimed were significant predictors of program success. Foremost among these were the designation of commercialization as an explicit program

objective with market stability and buyer awareness of the product. It was considered important that buyers, manufacturers, and other federal agencies were involved in project deployment, and that a "champion" of innovation be in a position of authority. Several factors were said to have a negative influence on the introduction of a technology, such as the OMB having a high level of influence, a lack of industry interest, rigid procurement requirements, high initial uncertainty in social/environmental and overall commercial risks, and excessive government regulation.

Baer, Johnson, and Merrow drew several important conclusions from their analysis of federally funded demonstration projects (1976), including the observations that a technology must be well in hand before moving to the technology demonstration stage, the necessity of relaxed time constraints, and the need for cost and risk sharing between the federal and local participants. Other key findings with possible relevance to the diffusion of solar energy technologies were that diffusion depends more on "market pull" than "technology push," that demonstration projects should have a narrow scope for effective use, that demonstration projects appear to be weak tools for overcoming institutional and organizational barriers to diffusion, and, on a more positive note, the dissemination of information from demonstration projects is generally not a serious problem.

The orientation towards the civilian consumer and the magnitude of the technology programs reviewed in both the McEachron and Baer studies suggest that their models of technology diffusion are relevant as models for the diffusion and later evaluation of solar energy technology programs, assuming, of course, the necessary adjustments can be made to tailor these models to the particular specifications of solar energy.

The history of the agricultural program has both positive and negative lessons to offer the developing solar energy program. On the positive side, the program has been successful for 70 years in achieving its goals of increasing agricultural productivity and raising the incomes of farmers. However, these very successes laid the germ for the socioeconomic revolution that has taken place in American agriculture and led to the corporate farm, thus undercutting the extension service's obligation to the small farmer. Again, similar to the analyses of federally funded technology development programs, the solar energy technology diffusion programs can learn from the agricultural extension programs' experiences, especially regarding the socioeconomic ramifications of the programs. Yet, it need be stressed that evaluation must be designed for the particular requirements of each program and the convenient extrapolation of solar energy evaluation from existing education, agricultural, or even technology diffusion models is problematic. Therefore, we should next inquire as to what forms of evaluation of solar energy programs are now underway and examine their appropriateness.

6.3 FEDERAL SOLAR PROGRAM EVALUATION

At the federal level, relatively little formal evaluation of solar energy programs currently exists. To the extent that existing solar programs are being evaluated, it is generally through informal, political processes. The Solar Lobby, an outgrowth of the May 1978 Sun Day activities, has acted as a critic of existing and planned federal solar energy activities. Another by-product of Sun Day, the President's Domestic Policy Review of Solar Energy (DPR), was intended to make some general solar policy recommendations to the President. The process of assembling the DPR resulted in the informal evaluation of existing federal solar programs, especially the public hearings in several locations across the country. These hearings provided the opportunity for interested members of the public to offer informal evaluation of current federal efforts in solar energy (Farhar et al. 1979). The General Accounting Office (GAO) has recently reported on the planning

and management of the ERDA solar energy program (U.S. Comptroller General 1976), the SERI and RSEC programs (U.S. Comptroller General 1977), and the effects of alternative funding levels on solar program performance (U.S. Comptroller General, February 1978). Whatever their merits, none of these studies was a formal program evaluation. A more recent GAO report on a program of solar heating demonstrations on federal buildings may qualify as a program evaluation, though the report's focus is on program management and legal compliance issues (U.S. Comptroller General, April 1978).

The OMB assesses all federal programs in the processes of developing budget proposals, but OMB interest is primarily management audit. Its concern for evaluation of program effectiveness generally is a function of how an evaluation is organized and its relevance to the program's continuation (i.e., funding level).

Federal agencies outside DOE (e.g., HUD) generally have their own evaluation procedures which may be applied to their solar energy activities. However, few formal evaluation studies of programs outside of DOE are known to be underway.

In DOE, principal responsibility for program evaluation resides in the Regulatory Program Evaluation (RPE), and Technical Program Evaluation (TPE) offices of the Policy and Evaluation (PE) division. The solar-related responsibilities of RPE include commercialization and state and local government programs; those of TPE include RD&D programs. The functions of these offices in relation to solar energy programs still are undefined. As noted in Section 2.0, DOE Memorandum 0203 established a broad requirement for four types of evaluation of DOE programs. Consistent with this directive, PE has not established strict department-wide evaluation requirements, but has moved carefully to determine the particular evaluation needs of specific programs and ways in which PE can coordinate and support evaluation activities.

A legislative requirement with possible significance for an internal evaluation of DOE solar energy programs is Title X of the Department of Energy Organization Act—the "sunset" provisions which require the President to submit to Congress by 15 January 1982, a "comprehensive review" of each DOE program for consideration in FY83 budget authorizations. Each program review is required to include an identification of the objectives intended for the program, the problem which the program was intended to address, and an assessment of the degree to which the original objectives of the program have been achieved employing methods of analysis appropriate to the character of the program. Title X also calls for an analysis regarding the program's beneficiaries and its impact on the national economy (e.g., effects on competition, economic stability, employment, unemployment, productivity, and price inflation). Finally, the program reviews are required to assess the degree to which the overall administration of the program has conformed to the objectives of the Congress in establishing the program, project the anticipated resources for accomplishing the unfulfilled objectives of the program, analyze the services which could be provided if the program were continued at a level less than, equal to, or greater than the existing level, and recommend necessary transitional requirements in the event that funding for such programs were discontinued. These requirements would nominally produce thorough program evaluations but similar clauses in the FEA enabling legislation did not result in any of the FEA programs being subjected to formal program evaluations or termination (House 1976).

The DOE Program and Project Management System adds little to the evaluation requirements of IMD 0203 other than to state that all program plans will be required to specify the criteria to be used for evaluating program effectiveness. This includes the technical and economic performance characteristics which constitute program success, milestones

and other quantitative factors which can serve as a measure of program effectiveness, and the probabilities of technical success of each discrete step of the R&D process (Meyers 1978). This information would certainly encourage, although not necessitate, formal program evaluation.

The Inspector General (IG) of DOE has important responsibilities for program review. However, as currently structured, these primarily concentrate on financial and management audit and compliance with legal requirements rather than evaluation of program effectiveness. Clearly, however, the purview of the IG could include program evaluation. The work of other agency IGs suggests that such evaluations could have significant impact on reviewed programs.

Some evaluations of the federal solar programs are in progress or at least being planned. The Treasury Department has plans for formal evaluation of the solar income tax credit recently passed by Congress, though details of the evaluation procedure are uncertain at this time. The HUD/DOE solar heating and cooling demonstration program includes a rigorous evaluation both of technical performance and commercial impact; the demonstration program presently is being reviewed by GAO, and also is included in a list of eight DOE technical programs under evaluation by PE-TPE. PE-RPE has let contract work in FY79 on regulatory, commercialization, and state/local programs. No specific solar energy programs are included but some of the programs to be evaluated may include solar projects.

The DOE Energy Extension Service (EES) has recently completed thorough evaluation of the 10-state pilot phase of the EES program. Although the EES program is explicitly concerned with energy conservation, many of the states' pilot programs have included solar energy projects as a component of energy conservation measures. The evaluation is intended to measure the ability of the states' pilot programs to achieve their own self-selected goals. The final evaluation report is based on correlations explaining the relationship among energy actions (or investments) taken and attitudes toward and knowledge about energy; the report has a heavy reliance on statistical analysis of variance of the six designated success measures. These data were reviewed with the staffs of the pilot programs before the final report was written to insure that the analysis is presented in an accurate context. As the first major formal evaluation of a DOE program, the EES evaluation study could set an important precedent for future evaluation studies in both the procedural and utility areas (see U.S. DOE, March 1979a).

6.3.1 Assessment of the Present Condition

There are several reasons for the present limited state of solar energy evaluation at the federal level. Among these are the newness of the federal DOE, the very diversity of its solar energy programs, the historic emphasis on R&D, a lack of guidance in organizing and conducting evaluation, uncertainty about who should perform evaluation, and the inherent difficulty of and ambiguity surrounding program evaluation.

Given that DOE is less than two years old, expecting highly developed evaluation processes for the federal solar program would be unrealistic. Building an organization with the scope and complexity of responsibilities of DOE should be and has been a slow, evolutionary process. Understandably, program building inevitably takes precedence over program evaluation in the investment of organizational resources. The demands of initiating programs leave program officials scant time to address the requirements of long-term evaluation. Furthermore, the diversity of federal energy programs, in solar energy as

well as otherwise, precludes a single, general-purpose model of program evaluation. Each of the solar energy programs poses a specific set of questions which, in turn, dictates a particular approach to evaluation. Useful evaluation is not achieved merely by administrative or legislative fiat, but by perspicacious design and execution of evaluation studies. The greater the variety of program types, the greater the time and effort required to develop applicable evaluation capacities.

Another reason for the present lack of formal evaluation of federal solar programs is the historic emphasis on R&D in the federal solar energy effort. The majority of federal solar program activity still is in the R&D area although this ratio is changing rapidly. The evaluation of technology RD&D presents particular evaluation problems which are difficult to handle. Still, past problems should not preclude future efforts, especially with the recent examples provided by McEachron (1978) and Baer et al. (1976).

Officials in ET, Conservation and Solar Applications (CSA), and other cognizant divisions of DOE involved in solar energy activities to date have had little formal guidance on the organization, performance, and use of program evaluation. The lack of strong direction and support for program evaluation throughout the department can be attributed to the exploratory state of evaluation within PE, which reflects the incremental development of the DOE. PE is an evolving division within a new and changing department; the federal solar energy program also is growing rapidly. To its credit, PE appears to be moving systematically to understand the requirements for evaluation within DOE and to develop the appropriate processes. Under these circumstances, to expect to have effective support and processes for solar program evaluation within less than several years is unrealistic. Thus the lack of guidance concerning evaluation is not only expected, it is probably beneficial at this stage of DOE's and solar technology programs' evolutions.

Another byproduct of the infancy of DOE is uncertainty about which office should perform evaluation of solar energy programs. Some ambiguity exists over whether the primary responsibility for program evaluation does or should lie with the program managers, with the Assistant Secretaries, with PE, with outside agencies such as GAO or Office of Management and Budget (OMB), or with independent contractors. Similarly, there is uncertainty about whether evaluation should be carried out mainly by in-house staff or by outside contractors, or by some combination of both. This ambiguity manifests the importance of support for evaluation by the top levels of the department. Because there is no one ideal way to organize and conduct evaluation, the present uncertainty could, in fact, be a benefit when specific program evaluation plans are formulated.

By this point, there should be little question that evaluation is a difficult task, both conceptually and practically. The inherent obstacles are magnified by the perception of officials of evaluations as threatening to their programmatic (and often professional) objectives. As Downs (1967) and Wilensky (1967), among others, have pointed out, the major criterion for organizational success in the public sector is the continued growth of an organization and its programs. There is little in the way of institutional counterbalance; program evaluation has an amorphous constituency which can disappear as readily as a given assistant secretary or the election of a new administration. Bureaucratic discomfort with program evaluation is not always unreasonable. For example, several DOE solar program officials have described their problem as too much, rather than too little evaluation. Program officials commonly say that they have inadequate resources to carry out a wide-ranging and complex program; in their view, resources allocated to program evaluation would be better invested in program implementation. They do not see program evaluation as an integral and priority part of program management. In addition, DOE program managers see the department as having had many external critics

(e.g., Congress, solar lobby) and few fans during its brief history. One consequence of these experiences is that DOE managers responsible for solar energy programs often tend to associate "evaluation" with undeserved criticism and attack. Given these perceptions, it is hardly surprising that public officials might greet program evaluation with skepticism or outright hostility. Whether such perceptions are justified is irrelevant, because they affect the prospects for useful evaluation of federal solar energy programs in either case. The solution to this problem almost certainly lies in reducing the anxiety about program evaluation by demonstrating that it can be a positive or beneficial element of solar policy management. The paradox of the situation is that practical experience with the benefits of program evaluation can reduce the anxiety presently attending program evaluation, but the anxiety is an obstacle to the acquisition of experience.

6.4 STATE SOLAR PROGRAM EVALUATION

The SERI research project on State Solar Energy Incentive Programs (see Ashworth et al. 1979) was a pilot analysis of the implementation experience of several types of solar energy programs in 12 states. The research was based on published information as well as extensive field interviews with state program officials, legislators, and grass roots activists. One objective of the study was to examine the evaluation practices and needs of the state solar energy programs. The remainder of this chapter is adopted from the final report of that project (Perelman in Ashworth et al. 1979: Chap. 7).

6.4.1 General Description of State Solar Program Evaluation

A complete examination of the evaluation practices in state solar programs is beyond the scope of this report. However, from the interviews and other studies of selected state solar programs that have been completed so far, several observations may be made regarding the current condition of state-solar program evaluation.

The principal methods of evaluating the effectiveness of existing state solar programs are the informal mechanisms. State legislatures exercise some oversight of established programs, but the practice is still inchoate, poorly understood, and uneven in most cases. For the most part, legislatures respond to external pressures of informal evaluation. When asked how the state legislature would determine whether specific solar programs are effective, one legislative analyst commented:

I think we'll find out [the] way the legislature finds out anything, and that is somebody coming in and complaining about the program not working right. [The legislators] don't usually go out and seek problems, but are willing to listen when someone comes in and says they have a problem.

The same analyst commented on the legislature's ability to conduct a more formal process of program evaluation.

We're getting more to that. I will not say yet that we really do have that kind of setup. The House's Natural Resource Committee and the Senate's Energy and Housing Committee have kept a pretty close watch on the work of the energy agency. There are several ways. One is by saying the energy agency will report back on specific tasks that are assigned to them by a certain date.

The executive branch view of evaluation was presented by a member of the staff of the agency responsible for implementation of the solar energy programs in the same state. When asked whether these programs need to be evaluated, the staff member remarked:

I think there always is [a need for evaluation], . . . because if we don't do it the legislature will come along and say, "You messed up the program," and blame the agency. [Evaluation is] really both the agency's and the legislature's role, but the agency really has to do it first. . . Evaluations, of course, get overlooked just because it's easier to go ahead with the program. Plus, it's hard to evaluate something that's within your own agency. But the agencies have to do that or else the legislature comes in and does it for you.

In general, the state agencies' motivation for evaluating their programs is essentially the same as the legislatures motivations; that is, in reaction to negative information from informal evaluators in the program's environment. The agencies do, however, try to develop some elementary measures of program activity (e.g., staff hired, grant applications received), particularly to bolster annual budget requests, which provides them with means and skills beyond the legislative staffs' capabilities.

State officials often lack a clear understanding of the meaning of formal program evaluation. It is not commonly distinguished from more informal evaluation and political processes, or from fiscal or management audits. This probably reflects the fact that program evaluation is still far from being a standard practice in most state government agencies. Simply put, many state officials have little experience with formal evaluation concepts.

In none of the states studied had formal responsibility for solar program evaluation been designated, nor was there any explicit legislative mandate for formal evaluation of solar programs. In one state, the major energy agency's office for program evaluation apparently disappeared, as a member of the agency staff explained:

[W]hen this [agency] was first formed there was one office called Program Evaluation, and all it did was critique the programs of everyone else. It disappeared. The person who was the head of it isn't fired or anything; in fact he has been promoted. . . . So I don't know why it did disappear. They did have some independent functioning. I guess probably if you would talk to the [head of the agency] he'd say that, yes, that office still exists really; it just isn't identified in our organization chart, and it's in my office.

In some state agencies, an office or individual may have the nominal responsibility for program evaluation, but true functional responsibility is vague or undefined. The lack of formal evaluation of state solar programs probably is not unusual. A 1972 survey of state program evaluation activities disclosed that only 21 of 42 responding states had any full-time staff explicitly assigned to doing program evaluation. Twenty-nine of the respondents characterized their program evaluation activities as inadequate; only two of the states provided any formal training in program evaluation (Hatry et al. 1973: 17).

The absence of any major effort to evaluate the effectiveness of state solar energy programs on the state level should not be viewed as unduly critical of those programs nor as a symptom of any lack of enthusiasm for the solar programs. Quite the contrary, the lack of serious demand for program evaluation is more an indicator of the broad base of popular support for solar energy and the eagerness with which state solar programs are

being initiated and developed. The emphasis in state solar programs is on near-term action, on expanding agency capabilities in a new area, and on meeting rapidly growing constituency demands for government support of solar energy. Under these circumstances, program evaluation inevitably receives lower priority in budgeting money, resources, and staff time. With little opposition to solar programs, no strong constituency for critical program evaluation exists. Whatever reservations agency staff may have about the ultimate effectiveness of the programs (and many had serious doubts about the effectiveness of specific programs), they perceive that they will be held accountable more for the level of program activity than for the effectiveness of the results, at least in the near term (see Clark 1975).

SERI interviews indicate that state officials almost uniformly think that formal evaluation of state solar programs is needed, and that the present level of evaluation activity is inadequate. However, there is considerable reluctance to dedicate a significant portion of scarce program funds and staff to program evaluations that seem unproductive in the near term. If program evaluation were funded by any means other than the program budget itself, state officials generally find it an attractive activity.

If and when formal evaluations of state solar programs are undertaken, state officials generally argued that they should be responsive to state government needs, but should be conducted by evaluators outside the line agencies. One state legislator's comments reflect the feelings of most state officials regarding the evaluation of state solar energy programs:

I do feel that if the state does [the evaluation] there is more consciousness or awareness of it, [and] more commitment to it. It seems to build a little more trust rather than have someone from the outside come in and do the evaluation. I think if it were done at the state level, it should be done by a third party. I don't want the state to evaluate its own programs. I would rather see a private agency or consultant evaluate the program . . . Maybe, on some ventures, it would be appropriate to have the [federal government do the evaluation], but somehow there really ought to be a state hook in there so the state takes ownership of its results, [so that] the state is committed to following it up. Speaking as a legislator . . . I would rather have some involvement earlier so that I, too, am aware of [the evaluation] and become committed to it . . . the legislative branch . . . has to be a little more explicitly involved along the way, not just at the end of it when the money requests come in.

The preceding section is not meant to imply that nothing is being done to evaluate state solar programs. Every agency studied is maintaining records and developing information to document the implementation of the solar programs. Generally, this is being done as part of the annual budget cycle to enable the agency to describe its activities and to justify its budget requests. The existing state solar energy program evaluations are commonly characterized by two recurring problems. First, the measures of program activity often are measures of program inputs or management processes rather than measures of program effects on the intended constituency. Second, even when the effects of the program are measured, the measures often fail to correspond to the objectives of the program. (These points are illustrated more extensively in the next section). In sum, the existing state solar program evaluation seldom generates the information that decision makers require regarding program effectiveness.

6.4.2 Status of State Evaluation by Program Area

This section reviews the present status of evaluation in each of the state program areas studied in this task: financial, RD&D, testing/certification, land use, and information/communication/education. The existing measures of program performance are compared with the major areas of concern for program effectiveness. These issues represent the types of questions that state evaluations of solar energy programs will need to address.

Financial. The most popular measures of the effect of financial incentives are the number of taxpayers claiming a credit, or similar accounting of the use of sales or property tax exemptions. Tax records also provide information about the persons claiming credits or exemptions. Some of the major areas of debate about the effectiveness of financial incentive programs concern questions of equity (e.g., do tax credits or financial incentives generally apply unfairly across socioeconomic classes?), legitimacy (are the incentives being applied to legitimate solar equipment?), the fiscal impact (what is the real cost of the financial incentives to the government?), and the potential diffusion effects (are the financial incentives actually stimulating the market for solar equipment and increasing the use of solar energy?). Generally, state agencies—like their federal counterparts—are measuring government inputs to the marketplace in the form of incentives. The outcome or effect of the inputs in terms of actual growth in solar energy use is much more difficult to evaluate.

RD&D. The main measures being used of RD&D program performance are the number and cost of research grants and contracts, the distribution of recipients of grants/contracts (e.g., universities, large versus small businesses), and tangible products (e.g., reports delivered, equipment, or buildings constructed). In some cases, there may be fairly sophisticated measurement of the technical performance characteristics of innovative systems. Those interviewed indicated some of the major areas of concern about state RD&D programs are the actual learning (e.g., are RD&D projects actually contributing new knowledge about solar energy use?), the value of the research (especially if the results of state RD&D activities relevant to state needs, resources, and opportunities), the transfer of results (are the benefits of RD&D activities being made available to potential users?), and the diffusion effects (are practical applications of solar energy being accelerated by the state RD&D projects?).

State solar RD&D projects are supported with the explicit goal of accelerating the use of solar energy in the state, yet few program measures have been established to demonstrate this goal. For example, in one state a major effort is underway to build houses using unorthodox construction materials and various types of solar systems. One program proponent indicated the program's ultimate objective is to change the standard practices in housing construction in the direction of greater energy conservation and major reliance on renewable resources. Further research disclosed that most of those involved in the program doubted the housing designs being demonstrated would be acceptable to the majority of the state's housing construction industry. Yet the evaluation measurements for this program focused on the technical performance of the innovative houses, not on the impact of the new designs on industry attitudes and practices.

Testing/Certification. The primary measures used to evaluate testing/certification programs appear to be (1) the existence of a certification program and (2) some measures of

compliance such as the number of participating manufacturers or the number of models tested and certified. Pertinent questions here include whether standards and compliance create a barrier to innovative solar technologies, if statewide standards are too inflexible to reflect local needs, resources, environmental conditions, and interests, and if state standards are effective in protecting consumers from illegitimate operators. Some state officials asked if standards and codes discourage low-cost, do-it-yourself systems so critical to decentralized applications such as solar technologies offer. Finally, others asked if standards were effective in protecting legitimate solar businesses from unscrupulous competitors and whether businesses need such protection.

Land Use. State land-use policies are particularly important because local planners, city councils, builders, real estate brokers, and citizen groups are vitally affected and, in turn, affect state land-use decisions. Conventional land-use practices are commonly viewed as a major barrier to solar energy use. Formal evaluation of innovative land-use procedures is needed to determine the effectiveness of new practices in facilitating solar development. Important areas of evaluation in the land-use area are the environmental, social, and economic impacts of solar-oriented land-use planning, protected access to sunlight, and the effects of solar land-use policies on the diffusion of solar energy. Measurement of the performance of land-use programs is limited; to the extent that any indicators are used, they seem to be confined to some evidence of the implementation of new planning activities and perhaps accounting for new developments using solar energy.

Information, Communication, Education Programs. Information, communication, and education programs may be the most prevalent forms of state government solar energy activity. They are also an area in which there is extensive experience and precedent for formal evaluation. Yet significant evaluation of these programs so far seems to be lacking. Common performance measures of these programs appear to be primarily input measures, such as funds and staff, number of materials produced, number of requests received, number of courses offered, number of meetings, workshops, and seminars held. Typical evaluation issues in these areas might be if the programs are satisfying the important needs for information and learning about solar energy. What education and training programs are necessary to meet the demands for personnel trained in solar energy technologies and how specific information and training activities affect the growth in solar energy use are questions commonly posed.

The Program Package. Formal evaluation of specific state solar energy programs by the states is still scarce. Equally important as the evaluation of these individual program areas is the evaluation of the total package of state solar programs (indeed, of the system of both state and federal programs). Programs need to be evaluated not only by themselves but in combination with each other, because in practice, they may conflict with one another or at least may fail to support each other's objectives. Tax incentives may be ineffective without equipment standards to determine eligibility; equipment standards may be inconsistent with building codes or land-use regulations; RD&D programs may promote technological innovation, while testing and certification programs may obstruct the penetration of innovative technology into the marketplace. It appears to date that no systematic evaluation is being made of the effectiveness of the total package of solar programs operating within a given state.

SERIO 

SECTION 7.0

DEVELOPING SOLAR ENERGY PROGRAM EVALUATION

This chapter addresses the specific needs, issues, and problems in developing solar energy program evaluation beyond its present condition. In particular, it considers the evaluation requirements for RD&D programs and poses some areas in which SERI's efforts might be particularly suitable.

7.1 TO EVALUATE OR NOT TO EVALUATE?

The principal question—whether to evaluate solar energy programs—has no simple answer. There are appropriate times and places for evaluation; for example, an early evaluation of a program might be held before it has had time to affect the situation or a later evaluation might come too late to reduce negative externalities. This section reviews some of the reasons for doing or not doing evaluation.

Good evaluation can lead to greater program effectiveness and the ability to meet the policy objectives. Presumably this is the major reason why evaluation is done. Given that solar energy is a new and rapidly developing field, well-designed evaluations should be especially helpful in improving the design and implementation of ongoing solar energy programs as well as future programs. Multiple evaluations of federal and state programs will be useful in assessing the decentralized nature of the many solar energy technologies.

Budget constraints will require priorities based on evaluation results. Rapid growth in solar program budgets cannot last indefinitely. As budgets become constrained, the problem of allocating limited resources among competing alternatives will require some indication of the effectiveness of existing activities. This also will help solar energy policy makers in comparing alternative methods of solar energy development.

Programs may be ineffective or even counterproductive. Ineffective solar programs waste limited resources. A real possibility exists that well-intended but ill-conceived solar energy programs may do more harm than good for the development and diffusion of solar technology. A recurring fear among solar advocates is that the acceleration of solar technologies, over optimistic cost projections, or faulty installation procedures could sour public acceptance of the solar energy technologies before they are sufficiently ready for distribution. One way to prevent these fears and wastes from materializing is the careful evaluation of such programs.

Policy makers need information about program performance and effects. Congress, the President, state governors and legislators, and local authorities are seemingly willing to expedite the development of solar energy but are uncertain how best to do so. The information gained from program evaluation are essential to reducing the uncertainties faced by these solar policy makers.

Programs easily can have unintended or unanticipated side effects that evaluation can help identify. This has two clear benefits. First, otherwise unrecognized benefits of the program may be revealed which may help justify the program's existence or expansion. Second, negative side effects can be discovered and relieved before they discredit the program or even create serious damage.

Finally, there may be political controversy over the design and management of existing programs which evaluation sometimes can help resolve by introducing empirical evidence into the discussion. Although analysis cannot totally replace political criteria, it can provide a common basis for discussion (see Siu 1978).

However, evaluation is not necessarily viewed as a universal benefit. Many people—especially program managers—have voiced a number of reservations regarding evaluation. For example, program managers may be embarrassed. Public officials should not be immune from embarrassment, of course, but solar programs are new and rapidly developing, which implies a great deal of trial and error. If program managers are publicly embarrassed too often by critical evaluations, they may become cautious, timid, and defensive at times when they should be inventive and risk-taking. Some argue that program managers need more help in program implementation rather than evaluation. A great deal of debate about the solar energy programs now is occurring which provides some evaluative data for decision makers. And program managers can justifiably claim that their resources would be better spent developing rather than evaluating programs. At both the federal and state levels, the demands on solar program officials seem to be growing faster than the available staff and other resources. Evaluation genuinely may be a lower priority need in the near-term than other program implementation needs.

Consensus about the formal goals and objectives of a program is not a requisite condition for evaluation. However, when the goals of a program are highly diverse, ambiguous, and in a state of rapid flux, it is difficult to design an evaluation which can respond to the key policy questions. Under such situations, it might be politically and methodologically prudent to wait for more stable conditions, assuming that delay is a feasible option, especially when a negative evaluation could tend to undermine program growth and development by creating uncertainty. Still, one can hardly oppose justifiable criticism; the key to establish an agreed set of criteria and produce a fair and balanced evaluation.

Finally (and again), there may be political controversy over the design and management of existing programs. Depending on the specific nature of the controversy and the actual situation, evaluation might be seized upon for purely political purposes. Evaluation may postpone the inevitable resolution of the conflict over the program, and, in the interim, merely prolong or even intensify the argument.

7.2 THE GENERAL CONTEXT OF SOLAR PROGRAM EVALUATION

The national solar energy program has several general characteristics that define the context and conditions in which program evaluation requirements should be developed. First, the solar program technology diffusion programs are both socially and hardware-oriented programs. The programs combine aspects of R&D with aspects of diffusion programs for technical and social innovations. This creates a need for evaluation theories and techniques which encompass both. Neither is an easy task; the combination is particularly difficult. Furthermore, the solar programs are characterized by the decentralization of their applications, both in their institutional structure and in its functional objectives. Although DOE has important lead responsibilities, many other federal agencies actively participate in solar energy programs; international agencies and state and local governments are also involved. Private business, grass-roots organizations, and the individual adopters play important roles in the distributed nature of solar energy resources. Most every community can make practical use of solar energy resources to which they have direct access.

The political constituency for solar energy development is diverse and broadening. The political interest in solar development has been growing rapidly in enthusiasm, as evidenced by the Sun Day activities in May 1978 and similar demonstrations in other nations. The widespread political support for solar energy has had two important consequences. First, for the past several years, program budgets have been growing at accelerated rates (see Fig. 5-2). The rapid growth of programs has minimized the current effective demand for solar energy program evaluation. Interest in evaluation of solar programs is favorable but weak relative to the more urgent concerns about establishing and expanding existing programs, and implementing new ones. Second, the explicit goals of solar energy programs have been vague and ambiguous because of the varied and often contradictory interests of the solar coalitions. There is little consensus among the solar energy constituency about how or why solar energy should be developed, only that it should be. This condition has several implications for program evaluation. The diversity of the solar energy development actors suggests a large number and variety of potential clients for evaluation, each with particular problems, whose goals, values, and expectations need to be reflected in evaluation criteria. As long as the constituency for solar energy programs is so variegated, the formal goals of programs will remain multiple and ambiguous, thus providing difficult guidance for evaluation criteria. This strongly suggests that the solar energy technology policy cannot be evaluated as an entity, that it must be disaggregated to the individual program levels to be effective.

7.3 STRATEGY FOR DEVELOPING SOLAR PROGRAM EVALUATION

If the decision is made to submit solar energy programs to formal evaluation, a few general requirements need to be considered in advancing solar program evaluations beyond their present status. A coherent strategy for developing solar program evaluation should include the anticipation of user demands, a developmental approach to evaluation, and efforts to integrate program evaluation into the overall policy process.

A danger inherent in the near future is the possible emergence of a crisis demand for evaluation in the solar program. A budget constriction in the solar programs could occur at the federal level within the next few years. The OMB has proposed a reduced DOE total budget for FY80, though the solar program budget is proposed to increase somewhat. Even if the President's Domestic Policy Review decision leads to an increased federal commitment to solar energy development—resulting in continued expansion of solar program budgets for several more years—the increasing budget competition between solar programs and other energy programs (or other government programs) is likely to stimulate greater questioning of solar program effectiveness, and thus an explicit demand for evaluation. Should solar energy program officials wait until the critical demand for solar program evaluation emerges full bore, they could find the delay in developing an evaluation capacity has left them lacking substantive evaluation results at the time when such evidence is most urgent. Building a formal, rigorous evaluation capacity only after the critical demand for evaluation has emerged means that needed evaluation results will not be available until after the crucial decisions about expanding, revising, or eliminating important programs have been made. The only other alternative is the "quick and dirty" evaluation whose methodologies and results are questionable; given the alternatives faced in this scenario, this option will almost certainly be entertained. Development of solar program evaluation capacity should, therefore, be initiated well in advance of demands for evaluation results which will probably develop in the future; indeed, program evaluation should be treated as an integral part of program planning and operation.

Anticipation of evaluation demands also means anticipating the needs of various types and levels of evaluation users. No single approach to evaluation will suit the needs of all potential clients. Fig. 7-1 is a suggestion of how different evaluation approaches might be tailored to the needs of different levels of the DOE solar energy program; similar formulation would be required for other federal and state agencies. For example, a White House level review would be directed at delineating broad policy options, such as DPR has done, while a department evaluation requirement would be much more program specific (e.g., level of funding or target clientele).

The second major requirement for developing solar program evaluation is the conceptual and methodological work on evaluation itself. A focused effort is needed on methodological research, evaluation needs assessment, program evaluability assessment, research on evaluation use, and staff training in evaluation for the overall area of solar energy program evaluation.

Finally, the utility of program evaluation is limited to the degree of its integration into the policy process. In the absence of a functional relationship to program selection, implementation, and termination, program evaluation will have little impact on government policy making. At the same time, the careful integration of program evaluation into the overall policy process will remind government policy makers that there is much more to policy making than the designation of a program; program managers will be able to see the beneficial results of program evaluation and perhaps be less opposed to it for their programs.

7.4 DEVELOPING EVALUATION BY PROGRAM AREA

In addition to the general needs for developing solar program evaluation, there are some needs unique to each program area. Some of the special problems and issues for evaluation of RD&D and commercialization program areas are considered below.

7.4.1 Evaluation of RD&D Programs

Evaluation of solar energy FCR&D programs means tracing various levels of both positive and negative outputs of such programs through the commercialization phase and into the mainstream of the society's technological infrastructure. This is a very circuitous, but hardly an impossible, enterprise (see Griliches 1979). The evaluations of FCR&D programs which have been done have focused more on the technology aspects of technology demonstrations and diffusions (McEachran et al. 1978; Baer et al. 1976); the social aspects of solar technology diffusion require a broader conceptualization of technology assessment and evaluation than has been the previous mode.

Rubenstein, who has extensively studied the problem of evaluating FCR&D programs, notes some of the complications in identifying the ultimate benefits and costs. For instance, government technical innovation and diffusion programs can rarely be restricted in their applications or monitored. The use of a particular process, product, material, or idea from a given development can easily appear in an entirely different area from that intended or envisioned by the innovators. The time lags are very long in commercialization and application of many R&D efforts and specific results become hard to trace. The RD&D paths of technical innovations are marked by modification, combination, and substitution, characteristics which are necessary and laudable but which render program monitoring much more difficult. Finally, and perhaps most intractably, there is lack of

<u>For Whom</u>	<u>What</u>	<u>By Whom</u>
Secretary Congress The Public	Broad program effectiveness; strategy; accountability, options	GAO; OMB
Assistant Secretary Deputy Assistant Secretary	Program performance; objectives; priorities in allocation	Consultants; AS Staff; OMB; GAO; CBO
Middle Management	Program performance and efficiency; allocation priorities; objectives	Staff; Consultants
Program Manager	Program performance and efficiency; budget priorities; tactics, milestones	Staff; Consultants

Figure 7-1. Solar Energy Evaluation Needs

an agreed-upon method of computing costs and benefits to various subsectors and society as a whole (Rubenstein 1977).

These problems make the evaluation of even the direct outputs of FCR&D programs difficult. The difficulty is compounded by the fact that the outputs of FCR&D are diffuse or "embedded" in many other inventions and innovations, as Rubenstein explains:

Many of the findings and innovative outputs of FCR&D are tightly embedded in products, processes, materials, and systems which are in wide and increasing use throughout society. . . . The difficulty with this embedded technology . . . is that it is not clearly discernible as a direct, integral unit contribution from a particular federal program or agency. . . . Any measurement system established for such "nonobvious" R&D outputs would have to be arbitrary and would have to probe deeply in a wide area of the public and private sectors. This does not mean that such an effort is not feasible or desirable. But it does mean that, if a serious effort is made to measure the results of R&D on a continuing basis, this important part of the output must be included and resources provided for the difficult job of detection and measurement. (Rubenstein 1977: 190)

In light of these problems, Rubenstein's overall assessment of the prospects for doing impact evaluation of FCR&D programs is important.

Despite the . . . difficulties, many attempts have been and are being made to do such measurement. I feel strongly [sic] that useful results can be achieved by improved methods of identification, measurement and analysis of the results of federal civilian-oriented R&D. However, in order to accomplish this, serious efforts will have to be made by both the originating agencies and other organizations to monitor and trace the flow of innovations from R&D to other "downstream" phases of the overall R&D/innovation process. That is, if there is a sincere interest on the part of the federal executive or legislative branches to determine the benefits from R&D on a continuing basis for the purposes of improvement in the process, legislation, and policy formation, then resources must be provided for the measurement and analysis required. . . . The entire U.S. R&D establishment has been relatively free of pressure for output measurement and evaluation until recently, and it will take some time for the people in the R&D/innovation process to get used to the idea that their output must and can be evaluated on a continuing basis and to tool up to contribute to it, so that it is not "done to them" entirely by outsiders. The idea of evaluating returns from R&D is not foreign to industrial R&D, where efforts have been made to do so for decades, at the individual company level. (Rubenstein 1977: 188-189)

Formal evaluation of FCR&D programs can produce useful information about the effectiveness of alternative RD&D processes (i.e., the design and implementation of projects, the use of contractors, staffing and management practices, or requisition procedures). Such evaluation of R&D programs can lead to development of generic guidelines for improving the productivity of broad types of FCR&D programs (e.g., Baer et al. 1976; McEachron et al. 1978). But for the response evaluation of FCR&D to be done well requires long-term, thorough tracking of RD&D program through multiple stages involving ultimate end-use applications and effects. The costs of such evaluation in terms of time, staff, documentation, and concentration of effort are considerable. Whether these costs are justified depends on the user's perception of the value of the resulting information.

The point should be emphasized. Impact, process, and response evaluations have distinctive requirements and serve different needs. To expect one to suffice for another is to expect too much. Again, the clients and requirements of the evaluation dictate the form it will assume.

7.4.2 Evaluation of Commercialization Programs

The evaluation of solar energy commercialization programs is straightforward; objectives are relatively well defined in terms of market penetration goals. The need for evaluation to support decision making in these program areas is well established, as is the mass of existing experience in evaluating most of these types of government activities.

Still, evaluating commercialization programs should not be seen as a trivial exercise. The major application of evaluation in the general commercialization end of the program spectrum is in the information, communication, education programs, financial programs, and government operation programs. The evaluation of demonstration programs in terms of measuring technical performance in field conditions is well defined but the evaluation of the broader objectives of demonstration programs in stimulating further development and acceptance of new technologies may pose some difficult evaluation problems in terms of objectives and time frames. The evaluation of regulatory and public works programs also is important but may be methodologically more difficult, stemming from the infeasibility of using experimental or even nonequivalent comparison group approaches. Time series data or less rigorous evaluation methods may be inadequate for addressing important policy questions in these programs. Indicative program measures, such as enabling legislation or general policy statements supporting solar energy development, usually are so diffuse in their effects that evaluation is extremely difficult.

The major problem in evaluating commercialization programs in general is that of scale. When full-scale commercialization programs are implemented without preliminary testing on a small-scale pilot, demonstration, or experimental basis, the prospects for evaluating program effectiveness are limited. The ubiquity of the new federal income tax credits for solar equipment makes statistical control or comparison of this program's effects methodologically difficult. State income tax credit programs provide a much better basis for comparing the effectiveness of tax incentives across states in actually stimulating solar diffusion, but as yet no formal evaluation of such programs is planned. An additional hazard of proceeding with full-scale national commercialization initiatives is not only that the effectiveness of such costly programs may remain uncertain but that the programs may be actually counterproductive. The possible negative effects of a full-scale, federal commercialization initiative may be very difficult to measure or correct until serious damage is done. The importance, therefore, of federal pilot programs (e.g., the EES program) and of the thorough evaluation of state and local solar commercialization efforts is difficult to emphasize, but even here the prognosis is mixed. The experience gained from pilot or demonstration facilities cannot always be easily extrapolated to full-scale programs.

7.5 SERI ROLE IN PROGRAM EVALUATION

The Solar Energy Research Institute (SERI) was created by the Solar Energy Research, Development, and Demonstration Act of 1974 with a broad mandate to promote the development of solar energy in the United States. SERI has six major operating divisions

for research, analysis, technology commercialization, information systems (including the Solar Energy Information Data Bank), international and academic programs, and program management (management of DOE contracts).

SERI is mandated to play a lead role in a comprehensive set of activities for promoting the development and dissemination of solar energy systems in the United States. But, at this time, SERI's institutional missions remain uncertain because of the newness of the Institute itself, the reorganization of its sponsoring organizations into DOE, and the normal tensions between two new organizations (Marshall 1979; Waroff 1979). A GAO report of September 1977 questioning the need for SERI (U.S. Comptroller General, September 1977) and Congressional creation of additional SERI-type installations, leading to the establishment of the four Regional Solar Energy Centers (RSECs), also have added to the obfuscation of SERI's institutional responsibilities. Because of these factors, some disagreement and ambiguity exists concerning the extent to which SERI is an RD&D organization, what role (if any) SERI has in solar energy commercialization, and the extent to which SERI's role includes policy analysis and other nonhardware-oriented functions, including program evaluation.

The relationship between SERI's institutional identity closely affects SERI's role in program evaluation. SERI's institutional status and client relationship will determine largely what, if anything, SERI can and should do in the area of solar energy program evaluation. Conversely, whatever SERI actually does in program evaluation will be an indicator of the Institute's functional identity and the receptivity of DOE program managers and officials to SERI's evaluation efforts.

Although the legislation creating SERI does not specify program evaluation, numerous references to evaluation, analysis, or program assessment appear in the legislative hearings. These references generally suggest some evaluative or critical role for SERI vis-a-vis ERDA or other federal agencies' programs. The ERDA request for proposals to manage SERI called for "a broad range of analytical studies and research tasks" to include "studying the nature, extent and objectives of federal participation" in solar energy development efforts, and "reviewing [the program] infrastructure and recommending . . . the need for different roles to increase effectiveness." In its contract proposal to organize SERI, the Midwest Research Institute indicated that the Institute would "develop a plan for continual review and evaluation of the national solar energy program," and "conduct the review and evaluation to examine the objectives, activities, nature accomplishments, and effectiveness of the program." And in its Mission Definition Report for FY78, SERI management defined SERI's program evaluation mission in terms of auditing national solar energy programs and providing thorough evaluations to ensure that public funds were utilized in an efficient manner. Even so, considerable uncertainty exists both within SERI and the DOE regarding SERI's role in evaluating other federal, state, and local, or international solar energy programs.

Within SERI, the concept of evaluation generally is based on the role of SERI as primarily an R&D laboratory. SERI staff are used as technical evaluators in solar energy science and technology. This kind of peer evaluation is a necessary activity, but one which has little relationship to the more comprehensive formal program evaluation processes and benefits discussed in this report. The danger of confusing R&D peer review with program evaluation is that it may divert attention from the important needs of formal evaluation of solar energy programs. Within DOE, SERI evaluation credentials are still not accepted for reasons already discussed in this report.

7.6 WHAT SHOULD SERI'S ROLE BE?

SERI was chartered by Congress to be a focal point for national solar energy program activity. Therefore, it is reasonable to ask what SERI's role is and might be in the evaluation of the great variety of solar energy programs found in federal, state, local, and international agencies because there are both advantages and disadvantages to an active SERI role in program evaluation.

7.6.1 Current SERI Activity in Program Evaluation

SERI has not yet initiated or conducted any formal program evaluation study. However, several planned tasks may lead to some initial evaluation work. This report represents the first SERI task to focus explicitly on program evaluation and is intended as a general primer in the concepts and methods of evaluation as they applied to solar energy programs. The pilot study of state solar energy incentive programs was conceived as a preliminary evaluability assessment and to develop initial experience and understanding of the analysis and evaluation needs of state government solar energy programs (see Perelman in Ashworth et al. 1979). A continuation of that task which would lead eventually to specific state program evaluation studies.

Another SERI task is to assist the Agency for International Development (AID) in developing procedures for planning and evaluating international solar energy programs in the developing nations. SERI's role in the National Plan for Accelerated Commercialization (NPAC) of Solar Energy, which was mandated by Congress, is currently under discussion; depending on what its role in NPAC is, SERI may be in a position to assist in the development of evaluation procedures for a wide range of federal solar energy commercialization activities.

7.6.2 Why SERI Should Do Program Evaluation

SERI has some mandate to do program evaluation. The mandate is more permissive rather than compulsive, but still provides SERI a justification for involvement.

The evaluator role is consistent with the concept of SERI as an independent leader of national solar energy activities. To the extent that SERI can be viewed as independent of the programs it might evaluate, SERI could bring the advantages of the expertise, impartiality, and perspective of the external evaluator to its evaluation work. Conversely, functioning effectively as an external evaluator would tend to reinforce SERI's analytic credibility to DOE and the solar community at large.

SERI has convenient access to both the technical and methodological expertise needed to support solar program evaluations. As noted earlier, solar program evaluators need not be experts in solar technology but occasionally will need to draw on such expertise in solar program evaluation studies. SERI's recognized leadership in solar energy science and technology offer a persuasive advantage to SERI's evaluation capabilities.

SERI may add visibility to evaluation results to the extent that it is recognized as a focal point of solar energy activity. Program evaluations from SERI may enjoy greater visibility and, therefore, may have better prospects for being used in correcting program deficiencies.

Last, SERI may provide leadership in ensuring that needed evaluation is done. SERI activity in the solar area may encourage the evaluation of key solar energy programs. Furthermore, SERI may be able to assist in securing funding for evaluation studies and could provide technical assistance in the design and performance of evaluations.

7.6.3 Why SERI Should Not Do Program Evaluation

A key argument against SERI doing program evaluation is that it lacks the necessary independence. SERI's reliance on DOE for funding and program approval may preclude a SERI role in evaluating DOE programs for at least two reasons. First, DOE program managers may attempt to prevent SERI from initiating evaluations or continuing evaluations in progress. DOE program managers may attempt to control evaluation results. In short, evaluating programs of managers who control SERI's budget may prove unfeasible. Even were these problems resolved, SERI's sole dependence on DOE may create an appearance of partiality that would a priori undermine the credibility of evaluation results. A second concern is that evaluation might conflict with SERI's program management responsibilities. SERI increasingly has been assuming the role of contract manager for many of DOE's solar program activities and may be unable to act impartially as an evaluator of programs for which it has major planning and managing responsibilities.

In addition to being dependent on DOE for its funding, SERI is mandated to promote actively the development and use of solar energy. Its mission then might be seen as biasing any evaluation results by being unwilling to be critical of efforts to develop solar energy. As noted above, credibility depends not only on the reality of objectivity but also on the appearance of objectivity.

SERI lacks extensive experience and resources in formal program evaluation. SERI staff have not yet conducted a formal program evaluation study. Building an effective evaluation capacity requires time and experience. Few members of the present staff have extensive training and experience in the design and execution of evaluation research, although several members of the staff do have relevant skills and experience in social research methods.

By the same token, the issue of a SERI role in evaluation depends on the comparative advantage of using SERI as opposed to other contractors. While SERI might have a comparative advantage to some divisions of DOE, several universities and independent consulting organizations are more experienced and better qualified to conduct evaluation studies than SERI. They may be relatively free of institutional bias or constraints and can be more flexible to respond to specific evaluation needs. SERI's absence from the evaluation lists does not necessarily limit the capacity for good solar program evaluation to be done.

And finally, the mandate and support for SERI to do program evaluation is uncertain. SERI has no clear obligatory mandate to do program evaluation, nor is there yet any strong client demand for SERI staff to perform evaluations, although the latter will probably develop, as argued earlier. Performance of program evaluation requires commitment and support at management or policy-making levels higher than those of the units evaluated, which, to date, has not been evinced. Program evaluation cannot be done without an unequivocal demand from policy and management policy makers at levels where the resulting information is intended to be used. With no client, there can be no evaluation.

7.6.4 Options for SERI Evaluation Role

In light of the above opportunities and constraints of SERI's evaluation activities, three guiding options for the SERI role in solar energy program evaluation can be noted: first, SERI could develop a formal evaluation capacity to serve client needs and generate its own evaluation studies; second, SERI could do little in formal program evaluation; and, third, SERI could develop a capacity for improving the general concepts and methodologies in the evaluation area.

The theses for the first two options are discussed above. The third option calls for SERI staff to perform few actual program evaluations and focus on the evaluation development tasks discussed above. This particular role for SERI in program evaluation balances some of the advantages and disadvantages of the first two options. The role of developing evaluation methodology is appropriately a centralized role of a single organization or group which can act as clearinghouse, broker, and technical advisor for the various actors in the field. Research and development of program evaluation methodology is a task suitable to SERI's R&D orientation. The minor disadvantage is that SERI's responsibility for solar energy research would bias the evaluation research toward solar energy evaluation, thereby limiting its general applicability or transferability to other areas and undermining the concept of evaluation R&D. A more serious drawback is that such an approach would not serve the immediate needs of the client.

SERIO 

SECTION 8.0

OBSERVATIONS AND RECOMMENDATIONS

A number of observations have been made in this report that bear repeating. First, solar energy RD&D programs are presently in a state of rapid evolution; different technology programs are situated at various stages on the development spectrum and are moving toward commercialization at different rates. Second, the U.S. Department of Energy, other federal agencies, and state agencies all have an interest and legal responsibility in the development and diffusion of solar energy. Their respective roles are still undergoing definition and coordination. Third, the development and diffusion arenas for solar energy are highly politicized. Finally, even under good conditions—such as program and organizational stability and the availability of key output measures—formal evaluation is not an easy task and rarely renders unequivocal findings. The concurrence of these observations makes it highly unlikely that the evaluation of solar energy technology programs is an easy venture, nor are the positive benefits often ascribed to evaluation studies guaranteed.

This is not to suggest that evaluation of the various solar energy technology programs from both their technological and social aspects should not be conducted. It is to warn, however, that the expectation of results should not be prematurely exaggerated. In many senses, DOE's reluctance to initiate formal evaluations of its solar energy program is well-advised. The programs are too diverse in natures, objectives, and end uses to fall easy prey to a single evaluation plan. Most interests would be better served if the nascent solar programs are permitted to operate until such time when objectives can be clearly defined, results can be observed and, just as important, evaluation plans can be carefully tailored to the specific requirements and goals of the programs being evaluated. Similarly, the solar energy program, because of the diversity of its components, cannot be evaluated as a single programmatic entity. It must be disaggregated if successful evaluation is to be performed.

These should not be taken as an excuse for evaluative inaction. As noted in the body of the report, the need for evaluation evidence will almost inevitably occur. To be asked for programmatic results when there has been no attempt at formal evaluation will almost certainly result in an evaluation which is conceptually and methodologically suspect, a position which casts grave doubts upon the evaluation results and leaves the program itself vulnerable. Perhaps the most useful evaluation exercise for many solar programs at this time would be for their staffs to begin prospective evaluation design exercises. Program planners and managers could develop evaluation designs as part of the program planning exercise that could be put into operation at the appropriate time; for example, objectives could be specified, data requirements defined, and analytic techniques explored. These evaluation design exercises would make the later formal evaluation itself more rigorous, reliable, and relevant to the policy maker and, therefore, more likely to be used. The experience of DOE's EES serves as a valuable precedent in this regard.

On a practical level, the understandable opposition of program managers to evaluation studies must be addressed. One partial remedy for this condition is the unequivocal support or formal program evaluation from the upper levels of management. But the key to overcoming this particular obstacle is the recognition that evaluation results are not necessarily devices to undermine bureaucratic domains or budgetary authorizations. On the contrary, sensitive evaluations can help the program manager in identifying those

areas of his or her program which are relatively weak and can benefit from special attention; similarly, it can identify those areas of strength which might be emphasized during program reviews. In the former case, this information is particularly valuable to the program manager, for it allows him or her to correct problematic areas before more critical external reviewers (e.g., the GAO, the OMB, or even an investigative journalist) discovers them and holds the program up to public ridicule. In either case, program managers need to understand that program evaluation can provide them with extremely valuable information that would improve their ability to manage the program. The key to this, of course, is that program managers be aware of what formal program evaluation can and cannot do and of their role in assuring that these objectives can be met; i.e., what can the program manager do to expedite the evaluation and assure its quality so that it serves his or her managerial purposes?

Program evaluation is not the exclusive responsibility of the program manager. The professional evaluation community has its role to play. In the case of solar energy RD&D, the conditions for evaluation are far from ideal. No convenient overarching model exists for the evaluator to use. Previous models of technology development and diffusion are sufficiently at odds with solar technology and diffusion processes so that their application is questionable, while the Baer and McEachron models require further verification before their relevance can be confidently determined. The evaluator must recognize and take into account the far-reaching technical, economic, political, and social ramifications of solar energy development and diffusion. The evaluator must also realize the problems inherent in operating in a highly contentious policy area and shape the evaluation to meet the particular needs of those operating in that arena. Finally, of course, the evaluator must be aware of the intricacies of solar technologies and how they affect the designated constituencies. Again, these features argue that the evaluation of solar technology development programs will scarcely be easy or straightforward, for reasons enumerated above, but it does not suggest that they are impossible.

More specific policy recommendations can be made. DOE must resolve its current ambiguity regarding evaluation studies. The EES evaluation and the evaluations currently being conducted by PE augur positively for the future but other issues need resolution. For example, will all DOE evaluations be conducted within the office of the Assistant Secretary for Policy and Evaluation or will the Energy Technology and the Conservation and Solar Applications divisions also have evaluation responsibilities? What are the mechanisms for assuring that evaluation results will have an impact on the program? And how will DOE regard evaluations made by external agencies, such as the GAO? On the surface, these seem to be little more than organizational issues that can be addressed by a simple department memorandum, but, in actuality, the resolution of such issues speaks eloquently to how DOE fundamentally views evaluation. A step in the right direction would be a departmental requirement that program evaluation plans—adjusted for the program's tenure—accompany budgetary requests.

State and local governments could have a crucial role to play in solar program evaluation. The variety and limited scale of state and local programs offer an invaluable opportunity for developing and testing government initiatives for solar energy commercialization. These programs need to be evaluated to provide essential information for federal policy decisions as well as to inform state and local governments themselves of the most effective use of their limited resources. It follows, then, that the federal government should provide some funding to initiate evaluation of state and local programs. The state and local governments should be prepared to invest some of their own resources in program evaluation as a means of securing their own involvement in and commitment to these evaluations.

The Solar Energy Research Institute can play a central role in evaluating solar energy RD&D programs. As noted above, SERI harbors the critical interdisciplinary skills requisite for effective evaluation; the combination of skilled and knowledgeable physical and social scientists on the SERI staff provides it with an unmatched comparable advantage. Furthermore, its physical removal from the daily operational requirements of Washington provides it with a detachment which is generally beneficial to evaluation efforts. Finally, its institutional image as an independent but sympathetic analytic organization might make it more acceptable to solar program managers than other external evaluators. For SERI to assume these roles, it must first persuade DOE program managers of the positive benefits accruing from program evaluation (even if they might be critical of a program), then demonstrate its capabilities to deliver a credible, timely, and relevant product. These capabilities can best be obtained and perfected by simultaneously evaluating ongoing solar energy programs and developing new methodological approaches to evaluation. The iterative process between evaluative theory and practice is essential if the currently posed problems with evaluation are to be overcome. Assuming these tasks can be met, SERI could indeed play a central role then in bringing the fruits of evaluation to bear in the solar energy RD&D.

Lastly, an underlying assumption to this analysis needs to be made explicit; namely, that evaluation is a useful—even necessary—part of the policy process. The difficulty of evaluation, especially solar energy program evaluation, has been reiterated throughout this report. The careful reader can well ask, "If it is so difficult, is it worth the candle?" Our answer is resonantly affirmative for two reasons. First, without evaluation, one could scarcely hope to improve ongoing programs, rectify program failures or negative externalities, or, in sum, serve the public interest in as efficient a manner as possible. Public service could hardly ask for less. Second, lacking evaluation, the integrated policy process would become a series of disjointed events. Program planners and managers would not be able to profit from past experiences, no learning experience would exist; one might as well develop, plan, and implement public policy in ways akin to the Queen of Heart's croquet match in which Alice watched the rules and equipment constantly being varied and the participants being threatened with the loss of their heads. Evaluation holds no guarantee for the strict adherence of public policy to its stated objectives or the rational workings of government, but it does offer a means by which programs can be measured and a learning experience can be obtained. These benefits are imperative if the solar energy technologies are to achieve their promises.

SERIO 

SECTION 9.0

REFERENCES

- Acland, Henry. 1979. "Are Randomized Experiments the Cadillacs of Design?" Policy Analysis. Vol. 5 (No. 2): pp. 223-242; Spring.
- Ashworth, John H. 1979. Renewable Energy Sources for the World's Poor. Golden, CO: Solar Energy Research Institute; May; SERI/TR-51-195.
- Ashworth, John H.; et al. 1979. The Implementation of State Solar Incentives: A Preliminary Assessment. Golden, CO: Solar Energy Research Institute; January; SERI/TR-51-159.
- Baer, Walter S.; Johnson, Leland L.; Merrow, Edward W. 1976. Analysis of Federally Funded Demonstration Projects: Final Report. Santa Monica, CA: The Rand Corporation; April; R-1926-DOC.
- _____. 1977. "Government-Sponsored Demonstrations of New Technology." Science. Vol. 196 (No. 4293): pp. 950-957; 27 May.
- Bardarch, Eugene A. 1974. "Subformal Warning Systems in the Species. Homo Politicus." Policy Sciences. Vol. 5 (No. 4): pp. 415-432; December.
- _____. 1977. The Implementation Game. Cambridge, MA: The MIT Press.
- Bauer, Raymond A., ed. 1968. Social Indicators. Cambridge, MA: The MIT Press.
- Bean, Alden, S.; Roessner, David J. 1978. "Assessing the Government Role in the Commercialization of Federally Funded R&D." Presented at the 1978 Annual Meeting of the American Chemical Society; Miami Beach, FL: September.
- Brewer, Garry D. 1974. "The Policy Sciences Emerge: To Nurture and Structure A Discipline." Policy Sciences. Vol. 5 (No. 3): pp. 239-244; September.
- _____. 1978. "Termination: Hard Choices—Harder Questions." Public Administration Review. Vol. 38 (No. 3): pp. 338-344; May/June 1978.
- Brewer, Garry D.; deLeon, Peter. The Scope of the Policy Sciences. Homewood, IL: Dorsey Press; forthcoming.
- Brock, Horace W.; Nesbitt, Dale M. 1977. Large-Scale Energy Planning Models: A Methodological Analysis. Menlo Park, CA: Stanford Research Institute.
- Campbell, Donald. 1969. "Reforms as Experiments." American Psychologist. Vol. 24 (No. 4): pp. 409-429; April. Reprinted in Caro, ed.; Readings in Evaluation Research (see below); Chap. 18; pp. 233-261.
- Campbell, Donald; Stanley, Julian. 1963. Experimental and Quasi-experimental Designs for Research. Chicago, IL: Rand-McNally.
- Francis Caro, ed. 1971. Readings in Evaluation Research. New York: Russell Sage Foundation.

- Carpenter, Polly; Hall, George. 1971. Case Studies in Education Performance Contracting. Santa Monica, CA: The Rand Corporation; December; R-900/1-HEW.
- Chelimsky, E. 1977. An Analysis of the Proceedings of a Symposium on the Use of Evaluation by Federal Agencies. Vol. II. McLean, VA: MITRE Corp.; July; M77-39.
- Clark, Robert. 1975. "Program Evaluation and the Commissioning Entity." Policy Sciences. Vol. 7 (No. 1): pp. 11-14; March.
- Clarren, S.; Woodward, J.; Hatry, H. 1966. Improving the Usefulness of Program Analysis and Evaluation Activities in Local Government: A Guidebook. Washington, D.C.: The Urban Institute; March.
- Coleman, James S.; et al. 1979. Equality of Educational Opportunity. Washington, D.C.: U.S. Office of Education.
- Commoner, Barry. 1979. The Politics of Energy. New York: Alfred P. Knopf.
- Costello, Dennis; et al. 1978. Photovoltaic Venture Analysis. Golden, CO: Solar Energy Research Institute; July; SERI/TR-52-040.
- deLeon, Peter. 1978. "Public Policy Termination: An End and a Beginning." Policy Analysis. Vol. 4 (No. 3): pp. 369-392; Summer.
- deLeon, Peter; McNown, Robert F. 1979. Potential Effects on the U.S. Economy and Society Resulting from a Dependence on Foreign Petroleum Sources. Golden, CO: Solar Energy Research Institute; February; SERI/TR-51-170.
- de Neufville, Judith I. 1978. "Validating Policy Indicators." Policy Sciences. Vol. 10 (No. 2/3): pp. 171-188; December.
- Downs, Anthony. 1967. Inside Bureaucracy. Boston, MA: Little, Brown, and Co.
- Drew, Elizabeth B. 1978. "HEW Grapples with PPBS." The Public Interest. No. 8: pp. 9-24; Summer.
- Elmore, Richard. 1979. "Organizational Models of Social Program Implementation." Public Policy. Vol. 26 (No. 2): pp. 185-228; Spring.
- Enthoven, Alain C.; Smith, Wayne K. 1972. How Much Is Enough? New York: Harper & Row.
- Farhar, Barbara C.; et al. 1979. Citizen Participation in the Domestic Policy Review of Solar Energy. Final Report. Golden, CO: Solar Energy Research Institute; January; SERI/TR-53-126.
- Festinger, Leon. 1957. A Theory of Cognitive Dissonance. Stanford, CA: Stanford University Press.
- Glennan, Thomas K. Jr. 1979. Personal Communication.
- Griliches, Zvi. 1979. "Issues in Assessing the Contribution of Research and Development to Productivity Growth." Bell Journal of Economics. Vol. 10 (No. 1): pp. 92-116; Spring.

- Hammond, Allan L.; Metz, William D. 1977. "Solar Energy Research: Making Solar after the Nuclear Model?" Science. Vol. 197: pp. 241-244; 15 July.
- Hargrove, Edwin C. 1975. The Missing Link: The Study of Implementation in Social Policy. Washington, D.C.: The Urban Institute.
- Hatry, Harry; et al. 1973. Practical Program Evaluation for State and Local Government Officials. Washington, D.C.: The Urban Institute.
- Hitch, Charles T.; McKean, Roland. 1960. The Economics of Defense in the Nuclear Age. Cambridge, MA: Harvard University Press.
- House, Karen E. 1976. "Getting Entrenched: Energy Agency Spends Much to Insure A Long Life, Foes Say." Wall Street Journal. pp. 1-19; 9 March.
- Janis, Irving L.; Mann, Leon. 1977. Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment. New York, NY: Free Press.
- Kline, David. 1978. Introduction to Educational and Social Research: A Modular Package Approach. Cambridge, MA: School of Education, Harvard University; Working Draft.
- Klitgaard, Robert K. 1978. "Identifying Exceptional Performers." Policy Analysis. Vol. 4 (No. 4): pp. 529-548; Fall.
- Lasswell, Harold D. 1971. A Pre-View of Policy Sciences. New York: American Elsevier.
- Levine, Robert A. 1970. The Poor Yet Need Not Have. Cambridge, MA: The MIT Press.
- Limaye, Delip R., ed. 1974. Energy Policy Evaluation. Lexington, MA: D.C. Heath.
- McEachron, Norman; et al. 1978. Management of Federal R&D for Non-federal Applications. Menlo Park, CA: SRI International; June.
- McLaughlin, Milbrey W. 1975. Evaluation and Reform: The Elementary and Secondary Education Act of 1965, Title I. Cambridge, MA: Ballinger.
- Marshall, Eliot. 1979. "The Solar Institute: Hobbled by DOE?" Science. Vol. 203 (No. 4386): pp. 1226-1228; 23 March.
- Meltsner, Arnold J. 1979. "Don't Slight Communication: Some Problems of Analytic Practice." Policy Analysis. Vol. 5 (No. 3): pp. 367-392; Summer.
- Meranto, Phillip. 1965. The Politics of Federal Aid to Education in 1965. Syracuse, NY: Syracuse University Press.
- Meyers, Dale, Under Secretary of the Department of Energy. 1978. Memorandum: "Program and Project Management System for DOE Outlay Programs." 31 May.
- Mosteller, Frederick; Moynihan, Daniel P., eds. 1972. On Equality of Educational Opportunity. New York: Vintage Books.

- Moynihan, Daniel P. 1973. The Politics of a Guaranteed Income. New York: Random House.
- National Science Foundation; National Aeronautics and Space Administration. 1972. Solar Energy as a National Energy Resource. Washington, D.C.: NSF/NASA Solar Energy Panel.
- Nay, Joseph; Kay, Margaret. 1978. Government Operations and Evaluability Assessment. Washington, D.C.: The Urban Institute; Working Draft.
- New York Times. 1979. p. D-8; August 13.
- Novick, David. 1975. Current Practices in Program Budgeting (PPBS). New York: Crane, Russak.
- Operations Research. 1971. Vol. 19: entire volume.
- Perelman, Lewis J., ed. 1978. Annual Review of Solar Energy. Golden, CO: Solar Energy Research Institute; November; SERI/TR-54-066.
- Perry, Robert L.; et al. 1977. Development and Commercialization of the Light Water Reactor. 1946-1976. Santa Monica, CA: The Rand Corporation; August; R-2180-NSF.
- Rivlin, Alice M. 1971. Systematic Thinking for Social Action. Washington, D.C.: The Brookings Institution.
- Roessner, David J.; et al. 1979. Application of Diffusion Research to Solar Energy Policy Issues. Golden, CO: Solar Energy Research Institute; March; SERI/TR-51-194.
- Rubenstein, Robert H. 1977. "Effectiveness of Federal Civilian-Oriented R&D Programs." Science and Technology Policy. Lexington, MA: Lexington Books; pp. 185-196.
- Rutman, Leonard., ed. 1977. Evaluation Research Methods: A Basic Guide. Beverly Hills, CA: Sage Publications.
- Scriven, Michael. 1967. "The Methodology of Evaluation." Curriculum Evaluation. Chicago, IL: Rand-McNally; AERA Monograph Series of Evaluation, No. 1.
- Searl, Hilton F., ed. 1973. Energy Modeling. Washington, D.C.: Resources for the Future.
- Sheldon, Eleanor B.; Freeman, Howard E. 1970. "Notes on Social Indicators: Promises and Potential." Policy Sciences. Vol. 1 (No. 1): pp. 97-112; Spring.
- Steinbruner, John D. 1974. A Cybernetic Theory of Decision. Princeton, NJ: Princeton University Press.
- Stobaugh, Robert; Yergin, Daniel, eds. 1978. "After the Second Shock: Pragmatic Energy Strategies." Foreign Affairs. Vol. 57 (No. 4): pp. 836-871; Spring.
- Stufflebeam, Daniel I.; et al., eds. Educational Evaluation and Decision Making. Itasca, IL: F. E. Peacock.

- Suchman, Edward A. 1967. Evaluative Research. New York: Russel Sage Foundation.
- _____. 1972. "Action for What? A Critique of Evaluation Research." Evaluating Action Programs. Boston, MA: Allyn & Bacon.
- Sui, Ralph G. H. 1978. "Management and the Art of Chinese Baseball." Sloan Management Review. Vol. 19 (No. 3): pp. 83-89; Spring.
- Twain, D. 1975. "Developing and Implementing a Research Strategy." Handbook of Evaluation Research. Vol. I. Beverly Hills, CA: Sage Publications.
- U.S. Comptroller General. 1972. Means for Increasing the Use of Defense Technology for Urgent Public Problems. Washington, D.C.: General Accounting Office; December.
- _____. 1976. Energy Research and Development Administration's (ERDA) Solar Energy Research, Development, and Demonstration Program Planning. Washington, D.C.: General Accounting Office; November; EMD-77-8.
- _____. 1977. The Solar Energy Research Institute. Washington, D.C.: General Accounting Office; September; EMD-77-67.
- _____. 1978. Evaluation Guidelines and Methodology: Program Plan for the Lead Division Area. Washington, D.C.: General Accounting Office.
- _____. February 1978. The Magnitude of the Federal Solar Energy Program and the Effects of Different Levels of Funding. Washington, D.C.: General Accounting Office; GAO.
- _____. April 1978. Solar Demonstrations on Federal Residences: Better Planning and Management Needed. Washington, D.C.: General Accounting Office; EMD-77-67.
- U.S. Department of Energy. February 1979. Domestic Policy Review of Solar Energy. Washington, D.C.: U.S. DOE; TID-22834.
- _____. 1979a. Evaluation of the Energy Extension Service Pilot Program. Washington, D.C.: U.S. DOE; March; DOE/CS-0075.
- _____. 1979b. Observations and Recommendations on the Future of the Energy Extension Service. Washington, D.C.: U.S. DOE; March; DOE/CS-0075.
- U.S. Executive Office of the President, Energy Study Group. 1969. Energy R&D and National Progress. Washington, D.C.: Government Printing Office.
- U.S. House of Representatives, Subcommittee on Energy and Power of the Committee on Interstate and Foreign Commerce. 1978. Land Energy Policies. Hearings conducted on 22 May, 5 June, and 9 June. Washington, D.C.: Government Printing Office.
- U.S. Office of Economic Opportunity. 1968. Instruction No. 72-8; 6 March.
- U.S. Office of Management and Budget. 1968. OMB Survey of Federal Evaluation Activities during FY 1977. Washington, D.C.: U.S. OMB; April; Bulletin No. 77-8.

- United States President's Commission on Materials Policy. 1952. Resources for Freedom. Washington, D.C.: Government Printing Office.
- United States Senate, Subcommittee on Intergovernmental Relations of the Committee on Governmental Affairs. 1977. The Sunset Act of 1977. Washington, D.C.: Government Printing Office.
- _____, Committee on Rules and Administration. 1978. The Program Reauthorization and Evaluation Act of 1978. Washington, D.C.: Government Printing Office; July; 95-981.
- van Manner, J. 1973. Program Evaluation: A Guide for Managers. Washington, D.C.: National Training and Development Service Press.
- Walling, V.; et al. 1979. "A Decision-Centered Policy Information Support System." Presented at ORSA/TIMS; May.
- Waroff, Deborah. 1979. "High Noon on the Sixth Day of Creation." The Energy Daily, Vol. 7 (No. 97): pp. 2-4; 22 May.
- Weiss, Carol H. 1977. "Research for Policy's Sake: The Enlightenment Function of Social Science Research." Policy Analysis, Vol. 3 (No. 4): pp. 531-546; Fall.
- Westinghouse Learning Corporation. 1969. The Impact of Head Start: An Evaluation of the Effect of Head Start on Children's Cognitive and Affective Development. Arlington, VA: Clearinghouse for Federal Scientific and Technical Information.
- Wholey, Joseph S.; et al. 1976. Federal Evaluation Policy: Analyzing the Effects of Public Programs. Washington, D.C.: The Urban Institute.
- Wildavsky, Aaron. 1974. The Politics of the Budgetary Process. Boston, MA: Little, Brown, and Co.
- Wilensky, Harold L. 1967. Organizational Intelligence. New York: Basic Books.

APPENDIX
CONSULTANTS

The following individuals were interviewed to gain their perceptions of program evaluations. Their comments served as valuable background for this report.

David Armor	The Rand Corporation
Bruno Augenstein	The Rand Corporation
Eva Baker	Director, Center for the Study of Evaluation University of California at Los Angeles
Victor Berlin	Experimental Technology Incentives Program National Bureau of Standards
Paul Berman	The Rand Corporation
Ed Blum	Office of Technical Program Evaluation Department of Energy
Monte Canfield, Jr.	Director, Energy and Minerals Division U.S. General Accounting Office
Launor Carter	Vice President, System Development Corporation
Sumner Clarren	The Urban Institute
Gregory Daneke	University of Michigan
David Dreyfuss	The Rand Corporation (Consultant)
Irwin Dubinsky	Office of State and Local Programs, Office of Regulatory Program Evaluation Department of Energy
Farrel Fenzel	Energy and Minerals Div. U.S. General Accounting Office
Howard Freeman	Director, Institute for Social Science Research Council University of California at Los Angeles
Ramon Garcia	Energy Extension Service Department of Energy
Carl Gawell	Office of Hon. Richard L. Ottinger U.S. House of Representatives
Thomas Glennan	The Rand Corporation

Lance Heiko	Planning and Assessment Division Northeast Solar Energy Center
Robert Hemphill	Deputy Assistant Secretary, Office of Technical Program Evaluation Department of Energy
Lawrence Kerpelman	ABT Associates
David Kline	Center for Studies in Education and Development Graduate School of Education, Harvard University
Dorothy Leonard Barton	Institute for Communication Research Stanford University
Richard Light	John F. Kennedy School of Government Harvard University
Michael McCabe	Staff Director, Environmental Study Conference U.S. House of Representatives
Normal McEachron	SRI International
Donald Nuse	ABT Associates
Maxine Rockhoff	Office of Technical Program Evaluation Department of Energy
Everett Rogers	Director, Institute for Communications Research Stanford University
Jack Roll	ABT Associates
Richard Rowberg	U.S. Office of Technology Assessment
Shalom Saar	Department of Urban and Policy Sciences State University of New York at Stony Brook
Al Schwartz	Energy Extension Service Department of Energy
Michael Scriven	Director, Evaluation Institute University of San Francisco
Hal Shear	Managing Director Contract Research Corporation
John Squitieri	The Rand Corporation
Victor Welling	Policy Systems Program SRI International
John Ware	The Rand Corporation

Document Control Page	1. SERI Report No. TR-51-294	2. NTIS Accession No.	3. Recipient's Accession No.
4. Title and Subtitle Solar Energy Program Evaluation: An Introduction		5. Publication Date September 1979	
7. Author(s)		6.	
9. Performing Organization Name and Address Solar Energy Research Institute 1617 Cole Boulevard Golden, Colorado 80401		8. Performing Organization Rept. No.	
		10. Project/Task/Work Unit No. 5328.40	
		11. Contract (C) or Grant (G) No. (C) (G)	
12. Sponsoring Organization Name and Address		13. Type of Report & Period Covered Technical Report	
		14.	
15. Supplementary Notes			
16. Abstract (Limit: 200 words) The evaluation of solar energy programs is an issue of increasing importance in the face of decreased energy supplies and the need to choose among competing energy technologies. The decentralized nature of some of the solar technologies and the possible social ramifications make the evaluations both more imperative and more difficult than might otherwise be the case. This report serves as an introduction to evaluation theory and methodology as they apply to solar energy programs; specifically, how formal evaluation procedures can be tailored to meet the particular requirements of the solar energy programs. The first part of the report provides a general introduction to evaluation theory and practice; this includes a discussion of evaluation typologies and methodologies in both abstract terms and in actual practice. The second section offers a general overview of the present condition of solar energy evaluation, assesses it in deficiencies in solar energy program evaluation, and poses some general observations as to how these might be alleviated. Included in these recommendations are a strategy for emphasizing the positive aspects of program evaluation and a specific evaluation role for the Solar Energy Research Institute.			
17. Document Analysis			
a. Descriptors			
b. Identifiers/Open-Ended Terms			
c. UC Categories 59, 60, 61, 62, 63, 64			
18. Availability Statement		19. No. of Pages 100	
		20. Price 6.00	