DOCUMENT RESUME

ED 187 269 HE 012 743

AUTHOR Feasley, Charles E.
TITLE Program Evaluation. AAHE-ERIC/Higher Education Research Report No. 2, 1980.
INSTITUTION American Association for Higher Education, Washington, D.C.; George Washington Univ., Washington, D.C. ERIC Clearinghouse on Higher Education.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 80
NOTE 68p.
AVAILABLE FROM Publications Department, American Association for Higher Education, Suite 780, One Dupont Circle, Washington, DC 20036 ($4.00)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Accountability; Administrative Policy; Budgeting; Case Studies; College Administration; Delivery Systems; Departments; Educational Policy; *Evaluation Criteria; *Evaluation Methods; *Higher Education; Models; Multicampus Colleges; Objectives; *Program Evaluation; Research; Self Evaluation (Groups); State Boards of Education; State Legislation; Statewide Planning
IDENTIFIERS Audits; Florida

ABSTRACT

The major distinctions between evaluation and research are examined, the chief differences being the intent and type of criteria against which judgments are made. Conceptualization of the evaluation process in higher education is discussed on two levels. A collection of nine similes for understanding evaluation is examined in terms of major activities, advantages, and disadvantages: evaluation as: (1) measurement; (2) expert judgment of worth; (3) assessment between performance and objectives; (4) the basis for decisions; (5) a goal-free process; (6) conflict resolution; (7) complacency reduction; (8) a change agent; and (9) ritual. Consideration is given to three types of formal evaluation models: the experimental, ecological, and eclectic approaches. The program evaluation process considered most restricting, that built into the budget process, is examined in detail. An investigation is also made of the purpose and practice of evaluation according to organizational level: state legislative audit, review by a state coordinating board, multicampus scrutiny, campus program evaluation, accrediting review, and departmental study. Factors that affect usefulness of program evaluation reports--such as timing, comparisons to similar units, and format--are discussed. The Florida State University system is described. Speculation about the future of the practice is made by examining present practices in diverse policy areas: nontraditional delivery systems, government revenue reduction schemes (taxpayer advocacy), management, regulation of professions, and consumer protection. (MSE)

PROGRAM EVALUATION

Program Evaluation

Charles E. Feasley

AAHE-ERIC/Higher Education
Research Report No. 2, 1980

Prepared by the
ERIC Clearinghouse on Higher Education
The George Washington University
Washington, D.C. 20036

Published by
the American Association for Higher Education
One Dupont Circle, Suite 780
Washington, D.C. 20036

This publication was prepared with funding from the National Institute of Education, U.S. Department of Health, Education, and Welfare under contract no. 400-77-0073. The opinions expressed in this report do not necessarily reflect the positions or policies of NIE or HEW.

Foreword

One aspect of "accountability" is to demonstrate that an activity or program is fulfilling its stated purposes. One result of decreasing income is to examine more closely the activities and programs of an organization to see if they are efficiently performed, effective, and needed. Thus the increased demand for accountability and the pressures of limited resources have greatly increased the need for program evaluation.

When program evaluation is suggested, it is often greeted with skepticism or apprehension. In its extreme, program evaluation is seen as a process to either legitimize an activity or to develop a rationale to cut back or eliminate. Obviously, there is a large middle ground for the use and results of program evaluation. Its success depends on how well the process is thought out, how accurately the data is gathered, and how honestly it is analyzed.

Before program evaluation can begin, there first must be some basic understanding concerning what is meant by evaluation and knowledge concerning the various evaluation procedures and techniques that are available. This research report by Charles E. Feasley, Coordinator of Operational Services, Extended Learning Institute, at Northern Virginia Community College, reviews and analyzes the major literature concerned with program evaluation. After discussing what is meant by program evaluation and describing nine ways that program evaluation is used, Dr. Feasley examines the various models that underlie all program evaluation activity. It is anticipated that this report will help to establish a more logical foundation for program evaluation and, when shared with all the parties involved, will help to develop plans of program evaluation that will encourage cooperation and minimize anxieties.

Jonathan D. Fife, Director

ERIC Clearinghouse on Higher Education

Contents

Overview

The major distinctions between evaluation and research are examined, the chief differences being the intent and type of criteria against which judgments are made. The intent of research is to know something in a generalizable way, while the intent of evaluation is to make a choice between options in a given situation. While both research and evaluation must have internal validity (measure what they are supposed to measure), external validity is much more important for research than evaluation. However, the data collected for evaluation must be considered believable by its users.

Conceptualization of the evaluation process is discussed in the report on two levels. A collection of nine similes for understanding evaluation is examined in terms of major activities, advantages, and disadvantages: evaluation as (1) measurement; (2) expert judgment of worth; (3) assessment between performance and objectives; (4) the basis for decisions; (5) a goal-free process; (6) conflict resolution; (7) complacency reduction; (8) a change agent; and (9) ritual. Both the evaluator and user of evaluation studies will expand their understandings of the milieu of the evaluation process by considering these diverse similes.

On a higher level of detail than the similes, consideration is next given to three types of formal evaluation models. The first model that emerges is the experimental approach. It is viewed as including goal-based, quasi-experimental, and classical experimental approaches. In general, the format assists policy-makers in deciding whether to continue or eliminate a given organization's program. It functions best when applied to uniform, fixed systems of activities and actors.

As questions mounted relative to a need for a process (formative) evaluation instead of only a product (summative) evaluation, interest shifted to an ecological approach, with its incorporation of multiple perspectives and methodologies. The ecological approach works best on a changing, complex program rather than a fixed one. Its aim is description and interpretation for program adjustment instead of prediction, which is thought to emerge from experimental approaches (Mims 1978).

Finally, as concerns arose about the subjectivity of ecological models, there emerged eclectic approaches that use features of both

experimental and ecological models. Clearly, it is important that an evaluator and consumer of each evaluation pinpoint all the purposes for a given study as a prelude to selecting among all the available evaluation approaches and techniques.

The most potentially program-constricting process for program evaluation is one that is a part of the regular budgeting process. Therefore, basic elements, benefits, and deficits of principal budgeting approaches are examined simultaneously with their ability to facilitate a quality program evaluation. Partial performance budgets and zero-base budgets appear to be more supportive of sound program evaluation practice than are incremental or formula budgets. However, much more time and line-staff involvement is required.

In this report an investigation is also made of the multiple perspectives on purpose and practice of program evaluation that exist according to institutional level: state legislative audit, review by state coordinating board for higher education, multicampus scrutiny, campus program evaluation, and department study. Evaluations were initially begun to determine the need for proposed programs and have spread to encompass most existing programs on a screening schedule if not an intensive basis. Another trend noted is that although early state-level reviews were limited to quantitative factors, more recent reviews have included qualitative considerations as well. The use of both kinds of data produces a more thorough, politically sound program evaluation.

Considerable debate has taken place on the extent to which goals and objectives should be the focus of evaluations. Deficiencies in the usefulness of program activity reviews has been related to a lack of precision existing in the statement and measurement of program outcomes. In response to such criticism, many schemata have been delineated for interrelating and valuing program goals and objectives. In this report discussion is also focused on the content of organizational goals and how to avoid the trap of examining only stated goals. Those actions to be taken because of these concerns include: (1) looking for unintended outcomes; (2) determining with program personnel what are realistic goals; and (3) measuring objectives at several points in the evaluation process.

The report states that the principal purposes of evaluation can be said to be planning, improvement, and justification. The phases of an evaluation process are seen to be foundation (establishing the scope of the evaluation), information gathering, and judgment. Participants in any evaluation process will be administrators and faculty

of the unit under evaluation, other institutional constituents outside the unit being evaluated, consultants from outside the institution, or a combination of any of these groups.

Factors that influence the usefulness of evaluation reports, such as timing, comparisons to similar units, and format are discussed at length. Multiple measures and diverse instruments are seen to improve the validity and reliability of a program evaluation. Also considered important are political and ethical aspects of conducting an evaluation.

Speculation about the future of program evaluations is made by examining present practices in diverse policy areas: nontraditional delivery systems, government revenue reduction schemes (taxpayer advocacy), management (measures of administrative quality), regulation of professions (quality assurance), and consumer protection (empowerment of the individual).

What Evaluation Is and Is Not

What We Treasure versus What We Measure

In a comprehensive study of the congruence between ideal goal preferences and goals actually observed by faculty, trustees, and administrators of a geographically dispersed set of institutions of six different kinds (public doctoral-granting institutions, private doctoral-granting institutions, public comprehensive universities and colleges, private comprehensive universities and colleges, liberal arts colleges, and two-year colleges and institutions), Romney and Micek (1977) noted that most of the measures of progress for goals *not* given a high priority by all three respondent groups are the data state agencies require institutions to collect. Some examples are the number of students enrolled, the number of full-time equivalent students, and grade-point averages.

As a result of diminishing state resources and the severe questioning of all societal institutions that has taken place in the past decades, there is a compelling need to measure what is transpiring within and as a result of society's social and educational programs. The two principal approaches to such measurement, evaluation and research, are often confused with one another. A discussion of the similarities and differences of these two fundamental concepts follows.

Evaluation versus Research

It is commonly stated that evaluation produces "worth labels" for some set of options within a decision situation. In contrast, it is said that research produces "truth claims" that are generalizable and serve as a basis for theory building (Popham 1975). Woodrow Clark (1977) suggests that it is useful to compare research and evaluation by looking at one's intention and one or more barriers to accomplishing that intention. Within research we want to know something in a generalizable way. The barriers to knowing in that way are twofold. First, the idea we want to understand has not been systematically examined before. Second, prior investigation of the ideas has been inconclusive. On the other hand, within evaluation we have the intention of making a choice between options. Barriers stem from the facts that the worth of the options is unclear and the information needs of the decision-maker are also unclear (Clark 1977, p. 9).

The most thorough discussion of similarities and differences between

research and evaluation can be found in Worthen and Sanders (1973). They focus their discussion on certain analytical elements:

Motivation of the Investigator: Research and evaluation are undertaken for different reasons. Research is undertaken in response to curiosity about an idea. Use of collected information is left to the natural processes of dissemination and application (Weiss 1972, p. 6), which is generally the publication of results. In contrast, because evaluation is intended to aid in the resolution of a particular kind of pragmatic problem, there is a very intentional scheme for distributing the results to decision-makers.

Objective of the Search: Different outcomes are sought by research and evaluation. An evaluation study will collect specific data wanted by a given decision-maker. Thus the study is directly requested and supported by that policy-maker to produce decisions. In contrast, only the investigators of research will select hypotheses and contexts in which to test hypotheses. The final outcome will be conclusions about phenomena.

Role of Theory: Research is the search for laws about the relationships among two or more *classes* of objects or activities, while evaluation describes the value characteristics of a specific thing.

Role of Explanation: A fully adequate evaluation can describe the value(s) of the subject under scrutiny without providing an explanation of how the effects are generated. On the other hand, research is conducted to pinpoint cause-and-effect relationships and trends.

Properties of the Phenomena: Evaluation is an activity designed to produce assessment of worth (social utility of a thing). Research is an activity intended to assess scientific truth, which is characterized by two principal properties: empirical verifiability and logical consistency.

Generalizability of the Phenomena: Evaluation can also be distinguished from research by the extent to which generalizations can be made about phenomena across time, across geography, and across types of educational activity. Research is said to produce results that are highly generalizable in all three ways. While product evaluation is usually generalizable with respect to geography, program evaluation has limited generalization with regard to all three elements (Worthen and Sanders 1973, p. 27).

Investigative Techniques: There has not been widespread agreement on the extent to which research and evaluation share investigative techniques. Several writers have stated that comparative experimental design, which is a frequent method of investigation within a research

study, is not appropriate for an evaluation study (Carroll 1965; Cronbach 1963; Guba and Stufflebeam 1968). On the opposite side of this issue, in addition to Worthen and Sanders (1973), those writers that have concluded that there is very little difference in investigative approaches are Stake and Denny (1969, p. 374), Weiss (1972, p. 6), and Hemphill (1969, p. 220).

Worthen and Sanders (1973) speculated that a researcher can "get away with" using the tools of only one discipline, since he chooses the question that is under investigation and would rarely ask a question that forces him outside the discipline in which he was trained (p. 135). Since the evaluator has to answer questions selected and defined by someone else, the answers are not as likely to be found by using only one standard methodology.

Criteria for Judging: Judging the adequacy of research and the adequacy of evaluation requires different criteria. Good research needs to have internal validity; that is, it should measure the effect of the variables under investigation rather than any extraneous influences. Good research should also possess external validity; its results should be applicable to other settings.

Primary evaluation criteria are isomorphism and credibility. The first criterion refers to the information that is gathered being in the same patterns as the information desired. The second criterion indicates that the information collected is believable to its users.

Both Weiss (1972) and Suchman (1967) have provided muted comparisons of evaluation and research by referring to "evaluation research." Suchman compares evaluation and evaluation research by saying that the latter comes closer to *proving* worth rather than asserting it (p. 8).

In summary, the major distinctions between evaluation and other types of research focus on intent and the existence of criteria against which one can make judgments.

Similes and Models

Within the literature of the emerging field of evaluation there have been two levels of concepts put forth to facilitate understanding. On one level are similes that explain evaluation by comparing it to other, more widely understood activities.

Evaluation as Measurement

The focus of this approach is on data and the formalized instruments used to collect them; frequent reference is made to standardized scales. Gardner sees the general process consisting of four steps: (1) identify attributes to be measured; (2) design and test an instrument; (3) use instrument under standard conditions; and (4) compare results to a standard (1975, p. 576).

Equating evaluation with measurement permits an evaluator to capitalize on the primary images of scientific measurement, those of reliability and objectivity. While measurement instruments produce data that are easily manipulatable into norms and standards, concern has been expressed about scores becoming ends in themselves, obscuring judgments and judgment criteria (Jemelka and Borich 1979, p. 264).

Failure of this model is likely if the entity to be evaluated does not possess significant measurable characteristics or the instrument used does not adequately measure the characteristic sought.

Measurement specialists such as Thorndike and Hagen (1961) concede that *true* evaluation involves the judgment of worth that exceeds the collection of measurement data:

> The term evaluation as we use it is closely related to measurement. It is in some respects more inclusive: including informal and intuitive judgments . . . saying what is desirable and good (p. 27).

The chief advantage of evaluation as measurement is the production of results that are comparable and replicable. On the negative side, the aspect of an entity that can be measured may be peripheral to the objective sought.

Evaluation as Professional Judgment

Common examples of the use of this evaluation model would be accrediting visitation teams, referees for journals, and peer review for awarding tenure or grants.

Criteria may be public. Methods include personal observation, interviews, tests, and review of documents.

The advantages of this approach are: ease of implementation; the inclusion of many qualitative and quantitative variables; and the quickness of the results and conclusions. The major difficulties have to do with undesirable subjectivity, potential unreliability, and difficulty of generalizing to other programs.

Popham suggests that there are two types of professional judgment approaches: those based on intrinsic criteria and those based on external criteria. To illustrate the difference between the two in his own way, Popham discusses the purchase of an electric drill. You could judge the drills on the basis of intrinsic characteristics: design, style, weight, and color (who wants an ugly electric drill?). You could also judge them on extrinsic factors such as how fast or how neatly they drill holes (who wants a glamorous drill that won't dent butter?) (1975, p. 11). Accreditation is a major example of professional judgment using intrinsic criteria.

Evaluation as Assessment Between Performance and Objectives

Popham (1975) labels this approach a goal-attainment model. Gardner suggests that no other type of evaluation has received more attention in recent higher education literature, encompassing as it does competency-based education and efforts to measure the program goal of equal opportunity for various subpopulations and education for coherent careers (1975, pp. 577-8).

Ralph Tyler (1950) is generally thought of as the father of the behavioral objectives movement. He advocated that the objectives of a program be spelled out in terms of specific student (client) behaviors. These behaviors are measured with either norm-referenced or criterion-referenced tests. The formulation of goals emerges from an analysis of three goal sources: the student, the society, and the subject matter.

A more recent goal attainment model has been proposed by Hammond, who discusses the nature of the institutional and instructional factors that might be relevant in the degree to which stated objectives are attained (1973). The steps in Hammond's model include: isolating the portion of the current program to be evaluated; defining the pertinent institutional and instructional variables; stating objectives in behavioral terms; assessing the behavior described in the objectives; and analyzing goal-attainment results.

Personal interaction with program staff is a common research tech-

nique of this approach. So is the examination of documents, process, and relationships. Both Scriven (1967) and Stake (1967) emphasize the early involvement of evaluators to assist in defining objectives.

A major criticism of evaluation as congruence between performance and objectives is that a focusing on measurable products rather than processes occurs. This may permit the over-looking of important side effects. Heavy emphasis is given to products that are student behaviors.

Evaluation as the Basis for Decisions

Although different language has been used by proponents of decision-oriented evaluation (Stufflebeam et al. 1971; Provus 1971; Alkin 1969), three basic assumptions were observed by Floden and Weiner (1976) as common elements to the individual works. First, stable publicly-proclaimed goals are the focal point for enacting programs. Second, evaluations collect information on the way in which programs function and on the effectiveness of programs in meeting stated goals (noting any discrepencies). Finally, it is stated that decision-makers will use evaluative information as a basis for program improvement efforts.

One major decision-oriented evaluation model has been developed by the Phi Delta Kappa National Study Committee on Evaluation: the CIPP (Context, Input, Process, Product) model. The CIPP model offers four types of evaluation activities: (1) context evaluation to help decision-makers determine objectives; (2) input evaluation to clarify ways that resources can be allocated to achieve project goals; (3) process evaluation to provide continual feedback to persons who must make various decisions during implementation; and (4) product evaluation to determine if an activity should be continued, revised, repeated, or concluded. The major steps in the process offered by the CIPP model are delineating, obtaining, and providing (Stufflebeam et al. 1971, p. 129).

Evaluation of this type is viewed as a continual exchange between evaluators and administrators. The methodology of evaluation is the methodology of an information system designed to provide information for project, program, and system decisions (Stufflebeam et al. 1971, p. 136).

A similar decision-oriented model has been advocated by the staff of the Center for the Study of Evaluation (CSE) at the University of California, Los Angeles (Alkin 1969). There are five stages of the CSE model. The initial stage is needs assessment, which consists of noting the difference between what is existing and what is desired. The principal focus of this stage is problem selection.

The second stage of the CSE model is program planning, with its major focus on program selection. The third stage focuses on program modification. It is known as implementation evaluation and provides information on the extent to which the program follows its own plan. The fourth stage, progress evaluation, also has a focus on program modification. This stage gathers information about how well the objectives of the program are being met. In this way products are examined enroute. The final stage is outcome evaluation. This refers to the collection of information about general worth of the program as reflected by the outcomes it produces. This final stage focuses on program certification, modification, elimination, or dissemination. The CSE model has been made more usable than several other approaches by extensive inservice training modules and workshops.

The principal criticism of the decision-oriented approach is that the evaluator accepts a decision context and values/criteria that have been defined by the decision-makers. Apple (1974) and Cooley and Lohnes (1976) observe that no evidence has been shown that the decision-maker has any more proficiency than the evaluator in these tasks of determining setting, program options, and priorities of worth. While Stufflebeam et al (1971) argue that an evaluator loses his objectivity (and usefulness) by participating in decision-making, Scriven sees an individual as abrogating his evaluator role when he fails to participate (1967).

Evaluation as a Goal-Free Process

In an attempt to avoid bias that may exist because of the narrow range of program developers' prespecified goals, an evaluator looks for *all* outcomes of a program including unexpected effects. These outcomes are examined and summarized in a single ranking of social utility.

Scriven brings to our attention methodological analogies of goal-free evaluation (1972, p. 3). In the field of aesthetics it is a common operating principle not to consider an artist's intentions in assessing a particular work of art. In philosophical ethics there has always been an argument between those who believe that the morality of an act is primarily determined by the motivation of the actor (he meant well) and those who would evaluate acts in terms of the consequences only. The double blind design employed in much scientific research is given as one further example.

The consumer of the program's services is seen as the major audience for goal free evaluation (House 1978). This approach is said to

be relentlessly comparative in nature (Scriven 1976). An evaluator may collect information relevant to program effects as they relate to accepted societal norms or some other type of generally recognized standard (Gardner 1975, p. 584).

Stake delineates a detailed process for conducting his form of goal-free evaluation: responsive evaluation. First, conferences are held with clients, staff, and audiences to identify program scope, discover purposes and concerns, conceptualize problems and identify data needs. Next, observers and judges examine selected antecedents, transactions, and outcomes. The third step is to prepare portrayals and case studies. Finally, reports are written and presentations are made (1975).

The goal-free evaluator must be skilled at interpersonal relations because of continued communication with program personnel. The lack of emphasis on formal measurement methods is said by critics to make goal-free evaluation too subjective.

Evaluation as Conflict Resolution

The commissioning of an evaluation study can be a signal that the program under scrutiny is subject to negotiation and modification. Individuals with opposing views will see the evaluations as a battle of worth, the outcome of which could determine a shift of program activities closer to their preferences (Floden and Weiner 1976, p. 8).

The major approach identified by the author as being consonant with this simile of "evaluation-as-conflict-resolution" is the judicial or adversary model of evaluation.

During the past decade considerable interest has arisen in adversary evaluations (Kourilsky 1973; Levine 1974; Owens 1973). The chief format used is that of the jury trial. The stages of the judicial evaluation model described by Wolf are typical. First is the issue-generation stage, in which a variety of persons involved in or affected by the program identify a broad range of concerns. The second stage, issue selection, centers on reducing the range of issues to a manageable number of the hearing. Next is the stage for the preparation of arguments, which consists of collecting testimony and abstracting relevant documents. The final stage is the hearing, with its presentation of arguments and panel deliberation (Wolf 1975a; 1975b).

Owens and Hiscox (1977) carried out six case studies of different uses of adversary evaluation and then compared them on the basis of purpose, format used, issue identification and selection, data collection for argument preparation, presentation, and decision-making. They noted three important spinoff effects of adversary evaluation:

better communication between evaluators and decision-makers; greater attention to formulating key evaluation issues; and increased concern for metaevaluation (the evaluation of the evaluation that has been conducted) (p. 20).

After participating in an adversary evaluation conducted by the Northwest Regional Educational Laboratory on Hawaii's 3-on-2 program, Popham and Carlson had six criticisms of the general adversary model: fallible judges; excessive confidence in the model's usefulness; disparity of adversary ability; potentiality for manipulating a particular result; excessive expense; and difficulty in framing the issue(s) (1977). Both Jackson (1977) and Thurston (1978) have observed that most of these criticisms apply to almost all other evaluation approaches. Both writers place great faith in the effect of the openness of the entire process. In at least one instance an outside anthropologist was employed to observe and record notes on proceedings; a videotape of hearings has been produced in many instances to secure reactions to the process itself (Owens and Hiscox 1977, p. 22).

In addition to the highly-valued characteristic of openness, a real plus is also found in the more active role that educational juries are encouraged to take. This may include questioning a witness and taking written notes during the trial. The process speeds discussion of pressing issues that might be debated in professional journals over a period of years (Jackson 1977).

On the other hand, among other unresolved problems of the adversary approach, the following have been listed: how confidential jury deliberations should be; the best working relationship for a jury composed of expert and nonexpert jurors; the size of a jury that is most efficient; the need for preexisting law (Denny 1976); when multiple hearings are justified; how "hard" data can be more effectively integrated with human testimony; and whether decisions must be made solely on the basis of evidence presented at the hearing (Owens and Hiscox 1977). Adversary evaluation is synthetic in allowing for the presentation and scrutiny of many different evaluation methodologies.

Beyond the frequent use of a jury, other adversary evaluations have employed debates and contrasting position papers. Thurston (1978) has explained the potential use of an appellate court model and an administrative hearing officer. The administrative hearing provides more public discussion and display than the appellate court, while the latter produces a written record, which is more useful in guiding future decision making (p. 6).

Evaluation as Complacency Reduction

The very act of participating in an evaluation may spur the consideration of new ideas by practitioners. Such participation produces a clarifying of program goals and available alternatives; furthermore, there may be increased satisfaction with the administration and evaluation of the program.

Despite the diverse purposes to be fulfilled by evaluation, a narrow range of evaluator skills (usually testing, survey analysis, proposal writing, and report production) has been favored thus far. To serve the role of evaluation-as-conflict-resolution or complacency reduction well, evaluators must have additional skills such as teasing out hidden goals and assumptions, training in mediation, and adeptness in interpersonal communication (Floden and Weiner 1974, p. 11).

Evaluation as Change Agent

"Formally documenting and describing what is already part of the informal communication network can have a powerful impetus for change" (Smock 1975, p. 4). If the evaluation is done by an "outsider" rather than an "insider," this documentation can have an even more profound effect.

Evaluation as Ritual

The conduct of an evaluation of a program produces a picture of government accountability and rationality that in turn promotes a feeling of security in taxpayers. The commencing of an evaluation suggests that the government is searching for improvements to its practices and solutions to existing problems. Also, evaluations may simplify complex social problems into a choice between clear alternatives. Smock (1975, p. 4) refers to evaluation as convention, standard approach, and liturgy. The ritual functions of evaluation are most strongly engendered when no recognition is given to that particular role of evaluation.

Formal Models for Allegiance

On a more formal level are detailed, multi-step models which describe the purpose and procedures for conducting a program evaluation. Classification of formal models of evaluation into experimental, ecological, and eclectic approaches has been made by Mims (1978).

Experimental approaches emerged from the natural sciences and psychology. They are used for accountability and specific decision-making. The chief models of this kind are the goal-based experimental

and quasi-experimental models, which have been described by Campbell and Stanley (1966) and Patton (1978).

Practical considerations that diminish the application of experimental approaches to evaluation are discussed by Stufflebeam et al. (1971): laboratory research designs require conditions that are difficult to achieve in evaluation settings; an evaluator must remain as unobtrusive as possible to program activities, rather than manipulating the environment as the experimenter does; experimental data collections occur at the conclusion of treatment and are of little value when correcting processes; experimental control requires the use of only one treatment at a time, a practice which is not possible with clients who may benefit from a treatment; and statistical techniques used with experimental procedures frequently offer restricted decision rules (a null hypothesis may be rejected or accepted, one treatment may be judged better than another), which may be inadequate for complex evaluation-based decisions.

While experimental approaches are quantitative, deductive, and uniform in nature, ecological methods (emanating from the disciplines of sociology and anthropology) are qualitative, inductive, interpretive, and diverse in nature.

The principal use of ecological approaches is to increase understanding and improve programs. Major models of this type include illuminative (Parlett and Hamilton 1976), transactional (Rippey 1973), and responsive evaluation (Stake 1975).

The major purposes of the illuminative model are description and interpretation. The process has been described by Partlett and Hamilton (1976) as: investigators observe, investigators inquire further, and investigators seek to explain. The techniques most commonly used are observation, interviews, questionnaires, and document analysis. The problem under investigation is said to define the methods, not vice versa. The audiences for these reports are program participants, program sponsors, and interested outsiders.

There are some problems with this approach. Its techniques are likely to be viewed as generally subjective, which is related to their difficulty of replication and transfer to other settings; and there is a considerable need for evaluators to have strong interpersonal skills.

Responsive evaluation is said to come closest to democratic pluralism (House 1978, p. 11). Any group that becomes actively involved in the evaluation process will have its concerns represented in that evaluation. Procedures are selected to fit the issues of interest.

Stake advocates the collection of descriptive and judgmental data

from various audiences. He further purports that the evaluator should be involved in the formulation of objectives as a program is planned. Problems are seen as best solved by local people.

Eclectic Approaches

Recently, several pragmatic models have emerged that combine elements of both experimental and ecological approaches. It is useful to examine one model in detail. Mims observes that eclectic approaches range from holistic evaluation (Rose and Nyre 1977) to decision-theoretic approaches (Edwards et al. 1975, 1978, p. 12).

The decision-theoretic approach has analyzed decision-making into four phases of activity. First, there is the recognition of a decision problem and the specification of its nature and scope. Second is the probability evaluation of likely posterior states of being. Bayesian statistics and techniques are viewed as particularly suitable for this phase. Such techniques permit quantitative combination of evidence from different sources and methods of inquiry (Edwards et al. 1975, p. 151). The third phase of the de ision-theoretic approach is ranking of outcomes according to desirability. Last is the actual choice among possible acts by using the values of outcomes and probabilities of states.

Evocation of Evaluation by Society

Those conditions identified by Peterson as predisposing factors to the implementation of state program review activities include: diminishing financial resources; declining public confidence in state government; state officials' stressing of accountability; and budget innovations by higher education agencies and institutions (1977, p. 10). Hill, Latterbie, and Stafford identified additional factors as: enrollment declines; the push for expansion in high-cost programs such as health-related professions; the slowdown in federal support; and the realignment of programs to meet standards set by the U.S. Department of Health, Education, and Welfare for an integrated system of higher education (1979, p. 1).

The turning point for plans to become programs is the authorizing cyclical budget. Any decisions to begin programs, modify their scope, or discontinue them (the major purposes of program evaluation) are most directly declared in the budgetary process and content.

Evaluation of Budgeting Approaches

After noting that a government budget serves three functions (controls spending, enables management of activities, and determines objectives), Schick suggested that the budgeting approach favored at a given time parallels the belief that one of these three purposes is being emphasized at the expense of the others (1971, p. 3). The first period of budget innovation occurred from about 1910 to 1935. This executive budget campaign was intended to emphasize control so as to prevent waste and corruption. Incremental budgeting is a common reminder today of that period.

The next era was that of performance budgeting, lasting until the late 1960's. The focus was on good management to achieve stated goals. Formula budgeting is one widespread artifact of this era. Recent interest has been shown in the use of performance budgeting for a portion of an institution's total budget.

The third budgeting movement focuses on planning. The development of planning, programming, and budgeting systems (PPBS) at the federal level represents the first event in the shift to planning budgeting. After use for several years, a number of states joined the

federal government in concluding that the time had not come for PPBS. One difficulty was that PPBS became a parallel and competitive process to the traditional budgeting approach, rather than supplanting it (Caruthers and Orwig 1979, p. 50). Zero-base budgeting is the latest approach to have an effect on the use of program evaluations.

Incremental Budgeting

Incremental budgeting is thought to be the most commonly used budgeting approach in higher education today (Adams, Hawkins, and Schroeder 1978, p. 54). Caruthers and Orwig see the incremental approach to budgeting as requiring the least work and analysis, while causing the least political conflict. It provides the least information concerning whether the budgetary decisions support institutional goals (p. 38). Hence, it has minimal usefulness in program evaluation.

Lingenfelter (1974) examined the operating budget requests, governor's recommendations, and final appropriations for higher education in Illinois, Michigan, and Wisconsin for the period of 1963-1974. He also interviewed more than 80 people in diverse decision-making roles before concluding that incremental budgeting models worked extremely well to explain appropriations outcomes.

In contrast, Bailey and O'Connor reanalyzed case studies collected by Wildavsky (1964), Fenno (1966), and Sharkansky (1968) to demonstrate the prevalence of incrementalism (1975). The surprising findings of the reanalysis was the significant number of instances of nonincremental changes in annual output at the federal and state level.

Formula Budgeting

In more concise characterization, Meisinger says that a formula budget is a combination of technical judgments and political agreements (1976, p. 2). This pair of elements is also found in program evaluations. Advantages and disadvantages of formula budgets can also apply to program evaluations.

These advantages have been attributed to formula budgeting: ease of preparation and understanding; clear use of financial incentives to support statewide priorities; equitable treatment of institutions (Caruthers and Orwig 1979, p. 43), preventing rich institutions from getting disproportionally richer (Moss and Gaither 1976, p. 553); prediction of future resource allocations; and making sure that higher education receives its share of total state resources, based on need and objective requests (Moss and Gaither 1976, p. 553).

Disadvantages noted for the use of formulas in budgeting include: failure to provide start-up costs for new or innovative programs (Meeth 1975); failure to react quickly to rapidly shifting price changes (Moss and Gaither 1976, p. 558); failure to measure more than one level of quality (leveling); and lack of flexibility to handle the complexities of enrollment decline (no consideration for economies of scale).

After concluding that 25 of the 50 states used budget formulas in allocating funds to higher education, a study by the Michigan Department of Education noted that if one considers quantitative guidelines as well as formulas in use by state governments, almost all state budgeting processes employ quantitative measures (1976, p. 15). Glenny et al. (1975, p. 46) refer to nonformula states as indicator states, where indicators are used to analyze but not generate budget requests.

Formula budgeting is not seen as being replaced in the future, only slightly altered in content to account for fixed costs not reduced by declining enrollments (Moss and Gaither 1976, p. 560; Caruthers and Orwig 1979, p. 45). The inclusion of qualitative elements is also seen for formula budgeting.

Partial Performance Funding

The Tennessee Performance Funding Project (TPFP) is an effort to improve the state's formula budgeting and appropriations process. It explores the question of whether it might be desirable and possible to allocate some portion of state funds to colleges and universities on a performance criterion (How good?) as compared to the current credit-hour and enrollment criteria (How much?) (Bogue 1976, p. 3). The underlying assumption has been that even imperfect measures, wisely chosen, may operationally improve the allocation process (p. 11).

The first phase of the TPFP consisted of involving national and state authorities in clarifying the conceptual base of the project, identifying related efforts underway around the country, outlining procedures for executing the project, and obtaining the necessary support for pilot projects (Bogue and Troutt 1977a, p. 4). The more widely used outcomes of this phase have been a set of hypothetical examples illustrating various approaches to performance funding (Bogue and Troutt 1977b) and a delineation of graduate competencies (Tennessee Higher Education Commission 1977).

Any modification in the funding policy was viewed as needing to

meet certain boundary conditions. They should: be politically acceptable, i.e., easily understood and accepted by legislators and members of state government; be professionally acceptable, i.e., striking the right balance between the need for institutional autonomy and the need for state-level review; encourage institutions to exercise initiative in developing performance measures on which they might eventually be funded; and recognize differences in institutional role and environment to promote diversity.

A description of the efforts of one institution during its pilot effort will clarify the scope and intensity of typical activities. Tennessee Technological University had a headcount of about 7,500 students at the time of the pilot project. It received $32,000 of external funds and spent about $25,000 of its own funds (not counting use of space and utilities) during the two years.

Faculty associates were selected to represent all five colleges and twelve selected departments of the university. After producing their own individual lists of instructional goals, a group list of 31 goals was then derived. Next, a survey was taken of all faculty (with almost a 90-percent return) as to whether each of the goals was presently a goal of that university and whether it should be a goal.

The faculty associates and project director then selected three kinds of performance indicators for each goal: (1) pre-existing data; (2) one or more extrainstitutional, standardized test; and (3) student and alumni reported satisfaction with personal goal attainment as determined by sample surveys (Dumont 1978a, p. 15).

Near the end of pilot projects, several conclusions were drawn about the various activities undertaken: many projects took advantage of existing good work on outcomes; the goals and indicators were chosen to reflect institutional diversity; and the goals and indicators show that public accountability concerns can be satisfied in distinctive ways (Bogue and Troutt 1978, p. 3).

In addition to those conclusions about institutional efforts, two altered visions of the context for performance funding also emerged. First, rather than awarding funds (usually one to three percent of the total funds for an institution) exclusively on the values of indicators per se, emphasis will be on the degree to which institutions publicize their performances. Second, instead of giving funds only as a reward for performance, it appears necessary to *stimulate* and reward performance measurement (Dumont 1978b, p. 20). An external panel of experts would determine whether an institution has satisfied its performance contract or not.

Several concerns remain about performance funding: the focusing of attention on indicators (the means) rather than goals (the ends); the uniform interpretation of performance data without regard to concerns such as reliability and validity; and the misleading comparisons of performance on similar or common indicators for institutions with differing missions, resources, and clients (Dumont 1978b, pp. 9-10).

Zero-Base Budgeting

In advocating the use of the current form of planning budgets, zero-base budgeting (ZBB), Pyhrr claims: that PPB focuses on what will be done, not on how to do it; and that PPB does not provide an operating tool for line managers who implement the policy and program decision (1973, p. 149).

Schick proposes that zero-base budgeting consists of three elements. First are the decision units of an organization, which have much to do with defining objectives and instituting sets of activities to accomplish stated objectives. Second, decision packages represent those sets of activities that are combined for the attainment of one or more objectives. Third, the managers of decision units and other higher-level administrators rank the importance of the various decision packages (1977). Thus, ZBB can induce integrated program evaluations at several levels.

Since zero-base budgeting was first used on the state-government level in Georgia, the observations of objective, yet close-at-hand researchers are worth consideration. Minmier and Hermanson (1976) surveyed budget analysts, then conducted follow-up interviews with selected analysts and department heads. While the majority of respondents felt that the quality of management information gathered under ZBB had improved, they did not believe that there was a significant reallocation of the state's financial expenditures. However, it did involve more line administrators in the budgeting process than earlier budgeting methods. Nevertheless, the first year of ZBB took substantially more time and effort than previous budget preparation.

Fincher noted that there appears to be no evidence that ZBB results in more clearly established goals or better measures of the progress toward stated goals (1978).

Other researchers have examined the implementation of ZBB in various locations to find problems and solutions that will generalize to subsequent adoptions and adaptations. Scheiring looked at the first-year use of ZBB in New Jersey and concluded that ZBB cannot be implemented overnight (1976). Not only do the proper forms have

to be prepared and staff trained, but strong support must be gathered from top management as well. A very recent survey of state budget officers by Ramsey and Hackbart pointed out the crucial role played by the governor and his executive administration as proponents and opponents to budget innovations (1979). All of these latter observations apply to the use and improvement of program-review processes.

Program Evaluation at Different Organizational Levels

Patton (1978) observed that there is no one effective strategy of evaluation in the abstract, separate from the organizational context in which it is introduced or the information-using capabilities of people employing it (p. 145). The key questions to be answered are Why? Who? and How? Mims focused on four broad purposes of program review: context, input, implementation, and outcome. Determining the need for a new program is a primary example of a context review. An input review collects suggestions on how a program should operate. An implementation (or management) review examines whether the program is being conducted as planned and with what degree of effectiveness and efficiency. An outcome (or impact) review not only asks Did it work? but probes any unanticipated outcomes and why they occurred (1978).

The possibilities for who should conduct a program evaluation could include: program personnel (a self-review); external reviewers within the institution (other faculty and other administrators); external reviewers outside the institution (accreditation teams, disciplinary specialists, professional evaluators); and multiple or mixed-group reviewers.

Because there are different participants and procedures used at the various organizational levels at which a program review may be commissioned (state legislature or governor, state coordinating board, multicampus system office, campus or department), it is vital that each of these perspectives of the program evaluation process be examined.

Legislative Reviews

Berdahl (1977) related details about the shift of the post-audit function for state programs from its early executive branch or independent status to a legislatively affiliated audit in over half the states (36 in 1975). Several states (California, Pennsylvania, and Wisconsin) have both executive and legislative auditors.

Pethel and Brown (1974) drew a distinction among three types of legislative audits or reviews. A financial audit examines whether money was spent according to legal procedures and for its allocated purpose. A management audit examines efficiency, the amount of resources needed to attain a particular program objective. Finally, a performance audit looks at the extent to which a program objective is met (ef-

fectiveness). For example, if the objective is to provide training for a particular population, its efficiency is that proportion of the population that received such training or that amount of the training activities individuals had an opportunity to complete. Performance audits are now mandated by sunset laws in over half of the states.

Sunset Laws

Sunset legislation was first conceived in 1975 and is now law in more than 29 states (Sherman 1978, p. 1). At least twelve states focus on regulatory activities. Nine state laws add advisory bodies and departments to the regulatory scope of sunset coverage. Seven states encompass all government agencies in their sunset laws.

Nineteen of the state laws require that preliminary reports must be written by existing government units. These reports will vary in quality according to the staff doing the reports and their objectivity. North Carolina established an independently-staffed commission to do the preliminary evaluations. However, most states use legislative audit, fiscal research, or substantive committee staff.

Although there has been considerable difficulty in collecting relevant data, sunset review has prompted agencies to organize information better and establish new data collection systems for future reviews (p. 18). The evaluation reports vary considerably in depth, ranging from New Mexico's 50-page report on 19 boards to a 140-page report in Tennessee on the Board of Accountancy.

Efforts to determine the costs of sunset evaluation studies have been unproductive because a considerable proportion of the expenses reflected start-up costs, and the early units that were reviewed had not been subject to too much legislative scrutiny previously.

At least two important lessons have been learned by evaluation staff about how to improve the sunset review process. First, all interested parties can be kept informed from the beginning of the review so that the adversarial nature of the process is minimized. Second, recommendations can be carefully offered so that they will apply to other agencies in addition to the one unit under scrutiny. Generally, there has been a good record of the acceptance of recommendations (p. 25).

Sunset reviews offer great potential for facilitating a dialog among legislators, administrators, and citizens. The process can realign the power position of the executive and legislative branches of government and can certainly increase the information available to decision-makers and the public.

Case Studies of Legislative Audits

Berdahl (1977) presents a study of the Wisconsin Legislative Audit Bureau interacting with the central administration of the University of Wisconsin over the request of the governor to prepare procedures to operate with reduced resources. Also examined by Berdahl is the 1973 evaluation of the Virginia College System by the Joint Legislative Audit and Review Commission.

Legislative Assistance

Since the objectives in higher education are not always known, are often in conflict with one another, and sometimes cannot be agreed on, the state-level review must be concerned with what ought to be done as much as with how to do it (Halstead 1974, p. 654).

A series of grants have enabled the Eagleton Institute of Politics to provide training and technical assistance to legislative staffs so that they can establish and/or improve program review processes. Such funding began in 1971 with a focus in six states on the legislative oversight of education. The most current grant supports the efforts of eight states in education and social sciences. One of the more important outcomes of these grants was a planning and implementing guidebook prepared by Murphy (1976).

State-Level Higher Education Agencies

Dougherty (1979a) described statewide reviews as a mechanism for providing local institutions with state, regional, or national perspectives (p. 11).

In his first comprehensive survey of state-level academic program reviews by higher education agencies, Barak (1975) looked at policies and procedures for examining new/expanded programs and existing programs. Seven major criteria were included in the state coordinating agency policies for approving new programs: program description, purposes and objectives, needs analysis, cost analysis, resource analysis, program accreditation, and availability of adequate student financial aid (p. 5). Barak also determined the extent to which states actually used quantitative criteria recommended by the Task Force on Coordination, Governance, and Structure of Postsecondary Education of the Education Commission of the States (1973) to guide program discontinuance. These criteria included: number of graduates in each of the past five years; number of students enrolled; size of classes; cost per program graduate; faculty workload; program quality as reflected by regional or national reputation; production of graduates from

similar programs in the state, region, or nation; economies and improvements in quality to be achieved by consolidation or elimination; student interest and demand; and appropriateness to institutional mission (p. 18).

Barak said that it is common practice for institutional reviews to focus on qualitative criteria, while state-level analysis gives almost exclusive attention to quantitative data. There is generally a two-phase process that occurs. In the first phase a screening process is used to identify programs that are of questionable need, productivity, quality, or other criteria (Barak and Berdahl 1977). A more intensive review is then conducted in phase two on the programs identified in phase one. Peterson (1977) noted that most states are now talking about qualitative as well as quantitative outcome measures (p. 3).

Berdahl (1977) described the differences between two types of performance audits: limited and intensive. He indicated that only a small number of states engage in intensive reviews. He provided the example of a comprehensive performance audit in Idaho that became so time-consuming that it has never been repeated on that scale again but rather has been replaced by management audits. The latter activity represents the kind of process, Berdahl noted, that has been advocated by the U.S. Government Accounting Office and others when faced with limited resources: evaluating the evaluations already being conducted by others.

Florida: A Case Study

A process for the systemwide review of *all* programs at graduate and undergraduate levels for selected disciplines at the nine state universities in Florida was begun by the Board of Regents in 1972. After having placed 16 graduate programs on probation for low productivity in 1973, the Board of Regents added 106 more programs to its probation list in 1974 and 39 more in 1975 (Florida Board of Regents n.d., p. 34). During this three-year probationary period the majority of academic programs either became more productive or were merged into similar degree programs. Eleven of the probationary programs were terminated by the university involved.

The kinds of data that were examined included the number of students enrolled, the number of students receiving degrees, admissions standards, faculty qualifications, curriculum and course offerings, budget, facilities, equipment, contracts and grants, library holdings, and placement of students (Hill, Lutterbie, and Stafford 1979, p. 2).

External consultants made site visits to campuses to gather data on

assess the scope of data that had already been collected by institutions themselves. The consultants were selected primarily from recommendations made by the faculties in the discipline under review, university administrators, professional organizations, and accrediting agencies. In addition to their own perspective of the discipline, they followed a set of guideline questions provided by the Board of Regents office encompassing the major categories of program quality, program priority, career implications, program administration and management, and articulation with other programs (Hill, Lutterbie, and Stafford 1979, p. 3).

Multicampus Reviews

In their initial study of multicampus systems, Lee and Bowen (1971) observed that reviews were conducted on individual campuses for proposed and new programs. Systemwide considerations were seldom included within such reviews. By their second study in 1975, the authors noted that systemwide reviews were a common practice for new programs, taking into consideration mission and academic quality. Also, in seven of the nine multicampus systems periodic reviews had been established for existing graduate and professional programs.

That actions by individual institutions are crucial is easily seen within Enarson's observation that it is anything but self-evident that state decisions, as opposed to local decisions, are made with a higher order of rationality and a clear adherence to the public interest (1975).

Institutional Reviews

Dougherty (1979) visited one private and nine public research universities that had reviewed and closed at least one program or that had undergone a serious financial crisis. He interviewed key administrators and faculty, examined written documents, and talked with appropriate state-agency representatives. Dougherty found that the authority to review programs existed at every possible level (departments, schools, total university or coordinating agency) in one university or another. Although a certain level may have the formal authority to review programs, it may be decentralized. For example, the coordinating and governing state agencies for higher education in Minnesota and Wisconsin, respectively, have allowed the main campuses of the primary state university to carry out program reviews (p. 8).

Dougherty relates the deficiencies that he finds in the various in-

stitutional program review processes to needs that can be met by quality institutional research (1979a). Financial analysis of the short- and long-run costs of program closure is very important, yet seldom done. Only one of the ten institutions studied had engaged in this activity, the private university. A closely related research activity is determining marginal costs when students are shifted from a closed program to another continuing program at the college. A third use of institutional research would be to cut down the time required to complete program evaluations. The primary example given to point out the potential magnitude of this program is the review of all graduate programs in the State of Washington which took nearly three years for the Council for Postsecondary Education to complete. By the time the review was concluded many program characteristics and personnel had changed.

A fourth use of institutional research is to generate comparable data. For instance, a statistic of faculty productivity (which is client contact) has great variation across disciplines.

While noting that the approaches discussed by Sprenger and Schultz (1974) and by the AAUP (1976) provided practical assistance to campuses facing reductions, Shirley and Volkwein found little intellectual rationale for actions (1978, p. 474). They synthesized many prior approaches to program assessment to derive a process for matching important elements of institutional mission; external needs, opportunities, and constraints; and internal strengths and capabilities. To facilitate comparisons, they selected evaluative criteria for quality (of faculty, students, library holdings, facilities and equipment); need (centrality to mission, present student demand, projected student demand, demand for graduates, locational advantages); and cost. Each criterion has its own rating scale. The comparisons place programs into five categories overall: (1) programs to be continued at the current level of activity regarding resource level, enrollment, and number of faculty; (2) programs to be continued at a reduced level of activity and resources; (3) programs in existence that will be developed further; (4) programs now in existence that will be phased out; and (5) new programs to be developed (p. 478).

A survey of 495 innovative institutions was conducted in 1975 by the Center for Research and Development in Higher Education at the University of California, Berkeley, to determine the use of evaluation practices (Hodgkinson, Hurst, Levine 1975). The response rate was 76 percent (375 institutions). Only slightly less than one-third of the institutions had an institution-wide committee on evaluation. In the

process of determining which standardized instruments were used in any of four evaluation areas (student characteristics, environmental measures, institutional goals, and course evaluation), there was a strong preference for locally developed instruments noted (p. 4). The analysis of purposes to which different types of institutions have put evaluation data showed no significant differences by type of control or level of highest offering. When asked about the most important evaluation problem with which they had to deal, 72 percent indicated that it was determining the effectiveness of new or existing programs (p. 3).

Several researchers have examined methods that have been employed by public research universities to review their existing programs (Braskamp 1979; Dougherty 1979b; Hall 1978; Shirley and Volkwein 1978). Regular committees or special task forces are the usual vehicles for conducting such reviews. Hall concluded that the principal benefits of the program reviews were increased information for participants and additional scrutiny of administrative decisions (1978).

Wood and Davis (1978) also summarize common methods of evaluating existing curricula: analyses of student transcripts, test of academic competencies, competency-based education of entire institutions, comprehensive examinations, examples of student work, institutional self-study instruments, surveys of current and former students, surveys of faculty, and program reviews.

At the same time that institutions and departments within them are subject to mandatory program reviews by state-level offices and agencies, they are also "voluntary" participants in program evaluations conducted by visitation teams from regional and professional accreditation associations.

Accrediting Reviews

From a content analysis of the published criteria for the six regional accrediting associations, Troutt (1979) identified five common criteria that were claimed by the accrediting groups to have some relationship with quality assurance: institutional purposes and objectives, educational programs, financial resources, faculty, and the library learning center. He could find no solid pattern or results from research studies to confirm or deny the claim that any of these characteristics represents a measure of institutional quality.

The extensive report by Harold Orlans et al. (1975) on private accreditation and public eligibility, ends with the conclusion that

neither private accrediting agencies nor government authorities are able and/or willing to control consumer fraud in education.

The Carnegie Council on Policy Studies in Higher Education (1979, p. 63) has recommended many changes in the practices of regional accrediting associations including: increasing the number of trained full-time staff to assist visiting teams, publishing periodically a report on the status of all schools that are members or have applied for membership, and making public final institutional evaluation reports.

Departmental Reviews

M. Clark (1977) surveyed 150 diverse and representative departments in colleges and universities on their program review practices. She found that 60 percent of the departments reviewed both their undergraduate and graduate programs. Objective statistics on various departmental characteristics (such as faculty training, experience, and publishing; number of degrees awarded per program; physical and financial resources; and student enrollment) were collected by 80 percent of the departments for internal use, which was less often than the same information was collected for outsiders. More subjective information (such as student evaluations of courses and teaching, faculty and student ratings of the departmental learning environment, and student judgments about their educational experiences) were collected more frequently for departmental use than for external use (Clark 1979, p. 2).

After demonstrating limitations of peer reviews for determining the quality of graduate departments, Clark (1974) shows the receptivity of graduate deans to a variety of multiple measures of quality. An exploratory study using departments of chemistry, history, and psychology at 25 universities across the country demonstrated that reliable judgments could be made from student, faculty, and alumni responses about program activities, procedures, and the learning environment (Clark, Hartnett, and Baird 1976).

A thorough description of the historical development and present implementation of the process for evaluating academic departments at the Urbana campus of the University of Illinois is provided by Smock and Hake (1977). The authors maintained that this systematic evaluation process differs from those elsewhere in that: (1) it is built on an extensive foundation of thought and planning going back a number of years, and (2) it is faculty based and largely a self-evaluative effort (p. 4). The coordinating group for this evaluation process is the Council on Program Evaluation (COPE), which is headed by an

Associate Vice Chancellor for Planning and Evaluation. A narrative report is written by faculty and administrators in the department following guidelines established by COPE. Discussed within the report are questions about the view of the discipline nationally, faculty research and service activities, operational procedures, and present problems. In addition, statistical data about tenure, promotion, courses taught and budget are provided by central university offices for inclusion in the report. The narrative report is submitted to COPE as raw data and is considered confidential. After discussion, COPE produces a public action report containing a summary of the self-evaluation, the recommendations made by COPE itself, and the reaction of the department to those recommendations released (p. 7). Since COPE sends its recommendations through normal administrative channels, its ideas get interpreted by individuals whose job it is to stay attuned to such external demands on the institution (p. 8).

Poulton (1978, p. 9) discusses the ways in which the usefulness of information provided by program reviews and the kinds of actions that are influenced by program reviews are different for various organizational levels: the department or program, the school or college, and the central administration. Single-unit reviews have the greatest usefulness for the unit examined. Benefits include improved procedures, clarified goals, improved internal communication, and improved rationale for resources. Problem areas are more readily recognized and approached in a rational manner.

At the college level, the information gathered from the review updates the existing knowledge of college staff, and permits defining current trends in the discipline. Generalizations about problems common to several units will slowly emerge. Reviews can be the basis for deans to reallocate funds and faculty.

The central administration receives the least direct benefit from a single program review. In rare instances it may result in major organizational changes or budget cuts. Usually it will allow top administrators an opportunity to observe the health of an academic unit as it responds to the review process and the recommendations arising from it.

Speculations by One Level of Review About Another

It is important to bear in mind the different values and procedures that are involved within program evaluation at different levels of organization. To obtain state and local agencies' views on federal educational program evaluations, the General Accounting Office

(Comptroller General 1977) sent questionnaires to state education agencies and a statistical sample of local school districts throughout the nation. Although state officials viewed federal managers as being most impressed by standardized norm-referenced test results, and local officials viewed state and federal officials in the same manner, state and local officials said that they are not most impressed by such results (p. v); in contrast, state officials said that they were most impressed by results from criterion-referenced tests, while local officials said that improvements in curriculum and gains in the affective domain were most valued.

Goals

Fincher (1978, p. 1) observed that institutions find it necessary to take an inventory of goals and objectives when there is a loss of sustained momentum or a failing sense of direction. As already discussed in an earlier chapter, institutions of higher education have been subjected to rigorous questioning of purpose during the past decade. The response has been the creation and use of schemata of goals and objectives such as developed by Peterson and Uhl (1977), Gross and Grambsch (1974), and Lenning et al. 1977. Such schemata are being used increasingly in program evaluations.

Conrad (1974) declared that in most universities, goals are often implicit, residing in an extended body of collective understanding rather than in explicit statements. Romney and Micek (1977) describe efforts to translate goals into measureable objectives. They note that a major difficulty in making this translation is the identification and agreement on pieces of evidence that demonstrate progress toward the achievement of established goals. As an example, Institutional Goal Inventory (IGI), an instrument developed by Educational Testing Service (Peterson and Uhl 1977), was used to compare the existing and ideal goals of faculty, trustees, and administrators at 45 colleges representing six major categories of institutions. The chief conclusion was that very little actual measurement of outcomes that the different groups felt ought to be measured took place.

Outcomes

Lenning (1977) provides a thorough analysis of extent configurations for measuring the outcomes of postsecondary education. He summarizes measures that affect individuals, institutions, society in general, and these groups simultaneously. He lists six attributes of an outcome. First is form: an outcome must be a product, an event, or

a condition. Second is the change status: whether the outcome preserves or alters the status quo of relationships and/or conditions. The essence of what is changed or maintained, known as focus, is the third attribute. The value neutrality of an outcome is the fourth attribute. The ease of measurement is the fifth attribute, and the final attribute is duration.

Five other factors boost understanding of outcomes. They are: which functional unit of the institution produces the outcomes; for whom the benefit is intended and who actually receives it; whether the outcome was intended; when and where the outcome occurred (p. 20).

After careful work in building a comprehensive outcomes structure, NCHEMS developed a series of standardized questionnaires to collect information from two-year and four-year college students in five categories: those just entering, those re-enrolling, those leaving without completing a program, those graduating or completing a program, and recent alumni (Gray et al. 1979). These instruments are already focal points for many program evaluations.

Social Indicators

The Department of Education in Oregon has a goal of setting its course of activities based on empirically verified needs. Without sufficient funds to expand student assessment efforts to gather that data, the use of indicators has been considered the most reasonable alternative (Impara 1977). Clemmer (1977) discussed the conceptual social indicator model that has been derived for the Oregon setting. Several types of indicators have been included in the Oregon approach: input, context, output (performance), and societal (side-effect). Based on a review of the literature of social indicators and analyses of the proposed use of indicators by the Oregon Department of Education, Jaegar suggested certain criteria for the selection or development of indicators. They should be: expressed in quantitative terms; time-referenced; directly and demonstrably related to a statewide goal; input indicators or context indicators, demonstrably related to at least one performance indicator or societal indicator; and accompanied by estimated measurement error for user enlightenment (Jaegar 1977, p. 22).

However, Jaegar declared the principal contributors to this area of research seem almost oblivious to problems of psychometric adequacy, although he does credit Land and Spilerman (1975) and De-Neufville (1975) with considering construct validity problems briefly (Jaegar 1977, p. 6).

Van Alstyne (1978) provides us with a perspective for seeing the use of social indicators at institutional and total system levels of post-secondary education. She notes that diversity and access are meaningful only when applied to the whole system of opportunities (p. 460).

Rossi and Gilmartin (1977) speculate that the future of social indicators will be characterized by a much wider group of data consumers. This is because the average person in future societies will have greater numerical abilities than is presently true. Another reason is that more government programs will require program evaluations.

Not Getting Trapped by Goals

Weiss has concisely stated that too much attention to the goals of a program can diminish the impact of an evaluation: among the many reasons for the negative pall of evaluation results is that studies have accepted bloated promises and political rhetoric as authentic programs goals (1973, p. 44).

Floden and Weiner (1976, p. 4) observed that because few goals have the strong support of a majority of citizens, the goals written into legislation will be vague enough to permit different constituencies to read in their own goals. Further difficulties in measurement arise because the outcomes of programs are obscured by the complex social context in which they occur, and public goals often change during an evaluation process.

Deutscher (1977) summarized previous research on organizational behavior, noting that organizations are rarely what they pretend to be by virtue of their stated goals. He referenced specific theoretical concepts that were developed for describing the evolution of goals, including goal displacement and goal succession.

Three ways to avoid the goal trap during program evaluation have been offered by Weiss (1973): (1) view success in terms of process, (2) be attentive to the unintended, and (3) negotiate a realistic scenario.

There are so many ways to conduct program evaluation that guidance is needed from respected practitioners about how to take the right action at the right time. In addition to the selection of formal evaluation models, decisions also have to be made about implementation of process elements such as Who? What? and Why?

Purposes

Lent (1974, p. 25) offers a useful trichotomy of purposes of program evaluation: program planning, program improvement, and program justification. Activities for program planning include an accurate diagnosis of client needs, the identification of program supporters and opponents, and delineation of means for attaining desired goals. Program improvement activities include comparisons with similar programs, improved communications among program participants, observations of whether strategies are working as planned, and conclusions about the adequacy of program responsiveness and flexibility. Activities viewed as falling under program justification may be listed as measuring the level of continued support for a program, discovering what supporters and opponents want to know about the program, demonstrating adherence to authorizing agreements, and advocating a future status (expansion, reduction, maintenance, or elimination of the program).

Phases of an Evaluation Process

Harshman (1979) examined many evaluation models or approaches taking note of similar component activities. Three phases of activities were observed. First is the foundation phase that forms the basis for subsequent evaluation actions. Activities included in this first phase would be identifying the decisions to be made, determining the goals and values of importance, and setting standards or establishing criteria. A second phase of activities, the information phase, consists of the specification, collection, analysis, and reporting of information. The judgments phase, which is the last, consists of three steps: a comparison of the reports from the information phase with the standards specified in the foundations phase; an analysis of how and why various elements combine to produce an effect; and making recommendations (p. 25).

Process Particulars

Mims (1978) provides an excellent discussion of options for designing a review process. In general, there are four types of groups for producing a program review process design: administrators, faculty, consultants, and a mixture of the other groups. A mixture group may be more useful and have greater support from the groups represented.

Three sources of ideas for the design process were described: adoption intact from other sources, modification of a process used elsewhere (adaptation), and specially created processes. While adoption of a process would result in the shortest time before implementation and would permit the use of a proven process, its results may be least fitted to the needs of the institutional constituents.

Mims also discussed several general characteristics of design processes. The design group can choose to operate in an open, consultative style or a relatively closed style. Some degree of openness is said to promote acceptance of the design and of the results (p. 5). All specifics of the review process can be spelled out prior to implementation or some details can be determined later. The former method of planning is termed a complete, comprehensive design, while the latter is called an emergent or phased approach. An evolving design process is thought to be appropriate when the institution is undergoing rapid change, when there is genuine uncertainty as to how to proceed, and when the institution is not highly experienced in program review (Mims 1978, p. 5).

Timing

The timing and form of evaluation activities depends very much on the purpose of the evaluation. The three major purposes of evaluation are represented by needs assessment, formative evaluation, and summative evaluation (Ball 1979). Needs assessment provides data for deciding whether to start a program. Formative evaluation is designed to feedback to program managers to aid in program improvement through modification. It usually occurs during the initial implementation of a program.

In contrast, summative evaluation is used to determine the overall worth of a program as the basis for expansion, maintenance, or extinction. It normally takes place after one complete cycle of program service has been completed and is frequently done by an outside evaluator. Stake offers a clever way to remember this sequence of evaluation: "When the cook tastes the soup, it is formative evaluation, and when the guests taste the soup it is summative" (1976, p. 19).

Although the formative/summative dichotomy has become almost a universal truth after its introduction by Scriven in 1967, Scriven has recently (1979) introduced the concept of preformative evaluation. Activities said to characterize this stage of evaluation are getting the evaluation budgeted and staffed, predicting probable program effects, and collecting baseline data.

Smith and Sanders (1975, p. 4) describe a predevelopmental phase of formative evaluation which includes the logical and empirical analysis of needs; in contrast, Bloom, Hastings, and Madaus (1971, p. 91) separate a diagnosis phase of activity from formative evaluation. Thus far, no common terminology is found universally.

Standing Committee versus Ad Hoc Task Force

Dressel (1976) pointed out some problems of using a standing committee. It must rely on the same channels of communication for gathering information as it has used for all other tasks. Its members establish interpersonal ties they feel obligated to defend. To avoid this latter difficulty, the State University of New York at Albany employed only external consultants to review its graduate programs (Mingle 1978). On the other hand, the Academic Vice Chancellor of the University of Illinois at Urbana carefully balanced disciplines and perspectives of education in choosing committee members (Braskamp 1979).

Comparative Approach

Scriven (1967) distinguishes between comparative and noncomparative evaluation, choosing the comparative orientation since decisions frequently have to be made among competing alternatives. He maintains that it is enough to identify which program can produce the greater effects without t[illegible] explaining why one program works better than the other. This [illegible] r task Scriven sees belonging to the educational researcher.

Whether it is program personnel or external evaluators that have begun to employ a comparative evaluation process, it is necessary to identify an appropriate domain of peer institutions. Increasingly, the techniques for making peer comparisons have become more systematic.

Recently, in employing a cluster analytic model for grouping peer institutions for research and admin[illegible]tive purposes, Terenzini et al. (1979) indicated improvements over earlier techniques for classifying institutions by reducing the arbitrariness and a priori specification of the classification structure, as well as the inability to accommodate

more than a few classification criteria. The authors cautioned readers about the need to relate aspirations to environmental constraints, while avoiding inappropriate levelling of quality. In a very formal way, both Michigan and New York have incorporated the peer institution into their state funding formulas (p. 21).

Measures of Use

Hutchinson describes three criteria for measuring the use of evaluative data by decision-makers. The first criterion, completeness, is the percentage of the decisions of a given decision-maker that are made to some extent with the use of the evaluation data. The decision-maker is asked to maintain a log of what information, if any, is used in making a decision. Focus, the second criterion, indicates the extent to which data were provided for the most important decisions. Hence, the decision-maker is asked to designate on the decision/data log which are the more important decisions. Finally, the third criterion, efficiency, refers to the percentage of the evaluation data used in making decisions. The use of predesigned data-gathering techniques can be the source of very poor efficiency of evaluation data use. Suchman (1967) stated that measures of efficiency arise from the examination of alternative program approaches in terms of costs: money, time, personnel, and public convenience.

Cost

Cost-effectiveness analysis emerges from the more general framework of cost-benefit analysis. The comparison of the monetary value of benefits with the monetary value of costs provides a measure for assessing the relative attractiveness of alternatives (Levin 1975, p. 6). A basic assumption for employing cost-benefit analysis is that the benefits of a given program can be valued by their market prices. Unfortunately, many social program outcomes do not have market values. In these many instances it is more important to relate costs to the actual physical or psychological outcomes rather than their monetary value (p. 9). Such a cost-effectiveness approach can be modified later into a cost-benefit analysis after experience demonstrates a market value for services and products.

One additional and similar approach to evaluation deserves mention. Cost utility analysis employs the decision-maker's subjective views in valuing the outcomes of alternative strategies (Fisher 1964, pp. 33-49). This technique is thought to be most useful where a complex set of outcomes is associated with each alternative course of action (Lif-

son 1968). Once the decision-maker determines the outcomes of alternative policy approaches and their utilities to him, these utilities are related to costs and probabilities of achieving the anticipated outcomes. The main objective becomes that of achieving the greatest utility within a fixed budget (cost).

Levin (1975) provides a discussion and examples of how to determine the costs of program components. He advises evaluators to expend effort to determine costs in proportion to the anticipated magnitude of contribution that the particular type of resource will make to the total cost of the program (p. 29). Levin also demonstrates that there are situations for which a marginal cost-effectiveness approach is most appropriate. For example, it would be valuable in helping a decision-maker choose between expanding an existing program or initiating another similar one (p. 50).

The speed with which one program rather than another produces desired results can be an important consideration (utility) for a decision-maker, even in preference to a long-term greater effectiveness by another (slower) program. An appropriate discounting of value (utility) in accordance with the preference for speed is required in such circumstances.

Sanders (1979, p. 14) reviewed seven current evaluation primers on the basis of 13 essential topics. Those receiving the lightest coverage were determining value, maintaining ethical standards, adjusting for external factors (political considerations), and evaluating evaluations. These situations and a few other pressing problems appear to deserve the additional discussion that follows.

Ethics and Standards

The planning for an evaluation should include an anticipation of potential conflicts among clients, evaluators, and audiences. As a standard practice there is a need to specify: (1) which existing program records will be examined; (2) what will be done with that collected data; and (3) who will have access to the completed evaluation report. Agreement should be reached on which sources of information are to remain anonymous and what safeguards will be taken to insure such confidentiality.

Understandings must also emerge on how much freedom evaluators have to collect information beyond that specifically requested by the contracting organization. Regardless of whether the latter condition is the subject of agreement, consensus is also needed on the extent to which outsiders can review the evaluation findings. Such a consensus may be part of a statement about the right of the client to terminate an evaluation process.

Anderson and Ball (1978) strongly advocate that evaluators should make their personal and professional value preferences clear. In the most useful way evaluators discuss how their biases are reflected in the choice of elements within the research design and implementation. As an aid to evaluators, Anderson and Ball have prepared a table (pp. 122-123) which shows how the design, measurement, analysis, and interpretation of an evaluation can be altered by an evaluator's preference for any one side of up to seven different bipolar ways of viewing the scope of an evaluation: (1) phenomenological versus behavioristic; (2) absolutist versus comparative; (3) independent versus dependent; (4) programmatic versus theoretical; (5) narrow scope versus broad scope; (6) high intensive versus low intensive; (7) process versus product. Also, professional value differences frequently emerge from

the different disciplinary backgrounds of evaluators. For instance, a sociologist may have preferences for context measurement as a primary method for conducting an evaluation, while a psychologist may wish to focus on interpersonal or individual development.

Anderson and Ball have prepared another table that spells out specific ethical responsibilities of both the evaluator and the commissioner of an evaluation during different phases of an evaluation process: making a contract; fulfilling the contract with as little interference of the program as possible; handling delicate extracontractual matters (unlawful practices, unsound activities); disseminating a balanced (objective) report; and guiding secondary evaluators (pp. 150-152).

In a background paper for the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Campbell and Cecil (1977) assert that research in program evaluation, social experimentation, social-indicator research, survey research, secondary analysis of research data, and statistical analysis of data from administrative records are and should be covered by Public Law 93-348 and other rights-of-subjects legislation. In such situations, they recommend the use of a conditional clearance affidavit in lieu of a full review by the appropriate institutional review board in most cases. Further, the authors suggest that changes in data collection procedures would be reviewed, not changes in administrative policy implementation. Also advocated by the authors is extending the right of informed consent into these areas of research, plus informing respondents of the risks of verificational interviews and subpoena of information where these risks exist (p. 21).

Another major focus of the ethics of program evaluation centers on whether there should be professional standards for evaluators. Prior to discussing specific qualifications under consideration across the country, it is important to take note of objections raised against making use of any standards *now*. Such objections have been summarized succinctly by the U.S. Government Accounting Office (1978). In the first place, standards may unduly restrict the supply of evaluators, driving up the cost of collecting evaluation data. Secondly, it has been hypothesized that the present time may be too early in the evolution of the evaluation profession to develop meaningful standards. In addition, an urging has been made that evaluation be viewed more like the journalism profession, which is seldom restricted, rather than the more regulated professions of law, medicine, and accounting. Finally, it is observed that standards for the sufficiency of evidence

may restrict an evaluator's capacity to provide data in time for use in a decision (p. 22).

Standards

In May 1974 the Joint Committee on Standards for Educational Evaluation began work on a volume to deal with issues and criteria for program and curriculum evaluation. This committee has 17 members representing a variety of educational organizations: the American Educational Research Association, American Psychological Association, National Council of Measurement in Education, American Personnel and Guidance Association, American Association for School Administrators, Education Commission of the States, Association for Supervision and Curriculum Development, American Federation of Teachers, National Association of Elementary School Principals, National Education Association, and the National School Boards Association. With a grant from the Lilly Endowment a nationwide panel of 36 experts was given a chance to discuss and write about various educational standards. Then a series of public hearings was held to permit wide opportunities for input and consideration.

Although the committee had initially planned to prepare both a detailed and a condensed version of the standards, work has stopped on the simplified version. The detailed version will probably include for each standard: a rationale, a list of guidelines for meeting the standard, a list of pitfalls commonly encountered by inexperienced evaluators in meeting the standard, and potential conflicts between the given standard and others. The four major areas of focus are accuracy, utility, propriety, and feasibility. The final standards should be published in 1980.

Brzezinski (1979) used the draft standards for educational evaluation in analyzing eight evaluation reports, which had been nominated to Division H of the American Educational Research Association for consideration in their 1978 evaluation awards competition. Her self-admitted cursory review showed that seven of the thirty standards were quite evident (six of the seven were from the accuracy category). The seventh prominent standard was balanced reporting from the proprietary category. References to an additional ten standards were discernible, mostly from the accuracy category, but there were three from the utility category and one from the feasibility category. Finally, she concluded that thirteen of the total thirty standards were *not* discernible in the evaluation reports examined. Four standards were from the utility category, seven were from the proprietary category

and two were from the feasibility category. Those reports displaying the greatest attention to educational standards were the longest and most technical (p. 5).

The Federal Joint Dissemination Review Panel (JDRP) has established 18 required considerations to determine quality educational products (Tallmadge 1977). Several of their primary considerations are defined by elements found in at least two of the areas of the Standards for Educational Evaluation. For instance, this is true for educational significance, generalizability, and credibility. However, after reviewing the criteria stipulated for endorsement by the JDRP, Hopkins (1977) expressed a strong reservation about their use. From an economic standpoint, it was thought unlikely that any sponsors would grant the time and funds needed to demonstrate satisfaction of all JDRP standards. Questioned also was the vagueness of certain standards (e.g. how professional the materials look).

Political Considerations

Anderson and Ball (1975) provide a concise definition of the politics of evaluation. It refers to the contest of power and wills in the evaluation setting, aimed at getting credit for success or avoiding blame for failure of current programs (p. 282). They observe that several common rationalizations may be given to diminish the value of doing an evaluation: the effects of the program are long range and cannot be measured well in a short period of time; the effects of the program are not measured well by existing instruments because they are very complicated; the impact and effectiveness of the program cannot be properly measured yet because the individuals most in need of it are not yet participating; and only those people who have been intensely involved in the program can estimate its impact.

Weiss (1975, p. 185) provides a vivid description of how different audiences await an evaluator's data as tactics in an ongoing political struggle. Politicians are concerned with satisfying constituents and keeping politically advantageous programs alive, whether or not they accomplish their stated goals.

In the first issue of the newsletter of the Evaluation Research Society, Cronbach (1977, p. 1) declares that evaluation is first and foremost a political activity. What is said to be needed is a political scientist to ask Laswellian questions: What is the motivation for setting up an evaluation? for agreeing to let evaluators collect data in one's school or community? for focusing upon, or ignoring, the evaluator's report?

Use of Evaluative Data

The extent to which information is used to make decisions that will improve program performance is negligible according to several organizational theorists (March and Simon 1958; Cyert and March 1963; Steinbruner 1974). They show that organizations search for new ideas and practices only when current performance falls below satisfactory levels. It must be further noted that the time required for evaluative information to influence decision-makers is much longer than usually suggested (Cohen and Garet 1975).

Poulton (1978) discusses factors in the organizational environment that influence the use of program reviews. First, he considers manageable factors. The credibility and perceived openness of the process are important. Credibility is enhanced by the use of top-quality people and recommendations that can be generalized to several settings or programs (p. 12). Internal communications to program staff members that convey fairness, candor, and flexibility boost the use of evaluations. The early delineation of what roles different people will play in the evaluation process maximizes an understanding of the extent of involvement and support for an evaluation.

Poulton also discussed factors that were not manipulable by the evaluator: how responsive given organizational units will be to participating in a program review, and whether the administrative climate is sufficiently stable and supportive to permit development and implementation of the program review.

The principal recommendations offered to guide evaluators in the preparation of useful reports focus on determining who the decision-makers will be, what information they will need, and when they will need that information. Morris and Fitz-Gibbon (1975) describe a process for figuring likely negative attitudes of the audience members and then adjusting the reporting process accordingly. For example, faculty members may have the attitude that special meetings are a drain on their valuable planning time. Recommended adjustments include putting the information in a brochure, on the bulletin board, or appearing only at a regularly scheduled faculty meeting (p. 30).

A major suggestion for preparing a presentation scheme is the use of multiple formats. In addition to a detailed report, leaflet-like summaries and oral presentations are also advocated.

Within the same report, encouragement is given to presenting data in diverse ways: verbal, numerical (tables), and graphical. When tables and graphs are to be employed they should be written prior to the narrative. Both tables and graphs should be designed to be self-

explanatory, since some readers may only look at them and not read the report.

Morris and Fitz-Gibbon (1975) provide readers with an entire chapter on how to prepare a variety of graphs and tables so that they are most easily understood by audiences. Particularly for an oral report, the authors recommend starting with simple and large visuals. In some cases the indication of trend data as shown by plus-and-minus signs may be easier to understand than actual data (p. 58). Tables are quite appropriate for showing the relationship among program components and the timeline for a project.

Several suggestions are made by Morris and Fitz-Gibbon (1975) on how to improve the readability of the evaluation report: define technical terms that are likely to be unfamiliar; use active verbs; use short sentences and paragraphs; and personalize the narrative whenever possible (p. 35). Popham advocated the inclusion of a verbatum transcript or anecdotal accounts of specific events to illustrate what activities are transpiring within the program (1975, p. 261).

The employment of adversary (contrasting) descriptions of the same program may help readers see more clearly the range of advantages and disadvantages of a given program than an "objective" view from one viewpoint. For such an approach, one staff member is given the responsibility to paint a positive picture while another staff member would paint a negative picture. The detrimental effects of having writers of unequal writing skills can be softened by having one person edit both viewpoints or one person write both the positive and negative accounts. This approach has been well exemplified by Stake and Gjerde (1975).

Providing program staff with an opportunity to offer a rejoinder to the evaluator's recommendations before they are made public, is the subject of considerable debate. For oral presentations, using a co-presenter from the agency being evaluated is a recommended technique in lieu of facing hostile audiences. Involving the audience in the presentation also lowers the resistance of people to new information.

Future Influences on Program Evaluation

A spatial configuration described by Windle (1978) has considerable merit as an aid in grasping the current divergent thinking about program evaluation and its likely destinations in the future. He views the scope of interests in program evaluation as three concentric circles. The smallest scope of interest is represented by viewing program evaluation as simply research to produce objective information others can apply. A larger sphere of interest is reflective of program values and those of its staff. Finally, the largest sphere of interest encompasses both smaller spheres of interest and also includes three societal interests in services: taxpayer advocacy, citizen advocacy (protection and problem solution are wanted), and consumer advocacy (fair treatment, which works toward independence).

Systematic Research about Techniques

Popham (1975) describes the critical need for systematic research about the effectiveness of various evaluation techniques. For instance, he notes that there is virtually no empirical evidence about the most effective needs-assessment approach or how to weigh criteria in reaching a summative judgment (p. 341). The amount of interference with ongoing program activities also needs to be part of the investigations of which techniques are most effective.

Gephart (1979, p. 2) recently declared that we now have a critical mass of people working on the explication of evaluation such that it no longer seems wildly optimistic to predict that work currently underway will merge by the start of the next decade to give us the conceptual and methodological clarity that has been so elusive.

Alternative Evaluative Criteria

To more fairly reflect the objectives and values of unconventional programs it is seen necessary to employ alternative criteria for evaluation. Gooler (1979) discusses one example, criteria that are particularly relevant to the nontraditional delivery approaches, which he labels distance education. Access is said to refer to how many people and what kinds of people have program resources available to them. A second criterion would measure the appropriateness of the program to the needs and expectations of target groups served. Gooler cautions

his reader that the relevancy is not static, since needs do change (p. 47). The quality of program offerings is a third criterion. Included within this criterion is the logic of the product and its scope. A fourth criterion consists of learner outcomes, both intended and unintended, including the attitudes that participants develop toward learning in general as well as toward the specific content covered. The impact of a program is the fifth criterion to be examined. It refers to the extent to which the program influences the mission, goals, and practices of other programs, institutions, or individuals. The sixth criterion is cost-effectiveness. It is important to measure the comparative cost-effectiveness of the program in relation to alternative uses of the same resources. Finally, the increase of our knowledge about the general field of delivering educational opportunities is also a criterion for evaluating such programs (p. 50).

Taxpayer Advocacy

Recent events in a number of states reflect increased taxpayer concern about the cost of government programs and their effectiveness. The shift of revenues for community colleges from local property taxes to state-level appropriations as a result of Proposition 13-type tax-relief measures may result in greater scrutiny of institutional programs by state officials. McCartan quotes the California Legislative Analyst as suggesing that Proposition-13 changes may mean the reevaluation of the state's policy toward oversight of the community colleges (p. 39). In fact, it is observed, the State Department of Finance, the Legislative Analyst, and the California Postsecondary Education Commission have endorsed the practice of annual reviews of community college budget requests.

The development and use of indicators of program administration represents a promising avenue for improving program evaluation. Sigelman (1976) suggested seven standards for evaluation of the quality of administration in the American state governments: professional quality as defined by (1) expertise, (2) information processing capacity, (3) innovativeness, (4) efficiency; and political quality as defined by (5) representativeness, (6) partisan neutrality, and (7) integrity. These standards were derived from the research literature on public and private organizations.

The embryonic nature of research in this area is demonstrated by the fact that Sigelman did not suggest any indicators for the standards of efficiency and integrity. For two of the standards, innovativeness and partisan neutrality, he suggested only one indicator measure. All

indicators that represent standards, including those multiple indicators for expertise, information processing capacity, and representativeness had not been examined for validity or reliability. Obviously, much research remains to be done.

Citizen Advocacy

Quality assurance is a process for providing citizen protection and reduced costs. It was first mandated by federal health-care legislation. Woy et al. (1978, p. 431) made a comparison of the contrasting emphases of program evaluation and quality assurance. Within the health context for which the authors write, quality assurance (in contrast to program evaluation) is generally patient/client specific, relies more on peer review (rather than administrative review), employs consensual methods (rather than empirical and normative approaches) to derive evaluative criteria, and has minimal aggregation of data commonly gathered through manual methods (rather than computer information systems).

In addition, the authors noted the gradual trend toward convergence of quality assurance and program evaluation. Key ways this would happen include sharing data, participation of evaluators in the peer review process, and the input of a variety of professional staff to the program evaluation process (pp. 440-441).

Consumer Advocacy

Cohen asserted that productivity in the human services has been achieved where service results in empowerment of the individual. This occurs when that consumer is able to (1) establish and achieve appropriate purposes, (2) clarify personal values and deal with value issues, (3) effectively understand him- or herself as well as others, (4) negotiate and work through the systems which affect his or her life, and (5) develop and use needed skills (1978, p. 38). A citizen empowerment chart has been prepared by Cohen to facilitate the measuring of progress toward self-sufficiency on the five dimensions of empowerment (pp. 40-41).

In an analysis of state government action, Jung et al. (1977) showed how state licensing laws and regulations contrasted with the Education Commission of the States' model legislation for approval of postsecondary educational institutions on a variety of facets: purpose, governance, and operation; course length, content, and objectives; degree requirements; staff qualifications; physical facilities; financial stability; public disclosure of materials; minimal qualifications for

entering students; recruiting practices; record-keeping practices; refund policies; and placement. In most cases, a majority of states did not have regulation as of January 1977.

In summary, it appears that there are many possibilities for student legal suits and/or federal and state intervention; however, concerted efforts to do program evaluation have been widely evident only during the past decade. Now there is that critical mass of people and ideas from many disciplines that is likely to forge worthy evaluation processes

Summary and Implications

To facilitate conceptual clarity, this report began by distinguishing between research and evaluation. Broad generalizability of data is characteristic of research, while immediate application to specific decisions is the overall purpose of evaluation.

Diverse ways of using evaluation can be considered by reflecting on the similes of evaluation discussed in this report. Both evaluators and those who request evaluation must be alert to the multiple purposes that can be attributed to an evaluation process. These individuals also need an understanding of the general approaches to conducting an evaluation (experimental, ecological, and eclectic) to achieve maximum comprehension and adaptability.

Since program evaluations have their greatest impact within the context of the budget process, an examination was made of the different roles various budget approaches play in program evaluation. Although considerably more effort and imagination are required to complete performance or zero-base budgets than incremental or formula budgets, the investment of time pays off in more suitable documents and processes for a quality program evaluation.

Another set of perspectives on the conduct of program evaluation corresponds to institutional level: department, campus, multicampus office, coordinating agency for higher education, and state legislative audit unit. Despite varying elements, common decisions about exact approaches can be found on each level; for instance, whether all programs will be evaluated rather quickly or a sample more intensively. Also the dominance of program goals over the evaluation process must be determined at each level. Varying amounts of encouragement can be given to the discovery of unanticipated outcomes in addition to the attainment of stated goals.

Procedures for conducting a program evaluation are selected to a large degree by determining the purpose of the evaluation. Form does follow function, whether it be needs assessment, program adjustment, or program verdict.

Superior evaluations result from attention to proper timing, the use of multiple measures and diverse instruments, and consideration for potential ethical dilemmas.

The future of program evaluation is said to be found in multidisciplinary efforts that serve consumers and citizens, as well as program personnel and legislators.

Bibliography

The ERIC Clearinghouse on Higher Education abstracts and indexes the current literature on higher education for the National Institute of Education's monthly bibliographic journal *Resources in Education (RIE)*. Most of these publications are available through the ERIC Document Reproduction Service (EDRS). Ordering number and price for publications cited in this bibliography that are available from EDRS have been included at the end of each citation. Readers who wish to order a publication should write to the ERIC Document Reproduction Service, Post Office Box 190, Arlington, Virginia 22210. When ordering, please specify the document number. Unless otherwise noted, documents are available in both microfiche (MF) and paper copy (PC).

Adams, Carl R.; Hankins, Russell L.; and Schroeder, Roger G. *A Study of Cost Analysis in Higher Education*. Washington, D.C.: American Council on Education, 1978.

Alkin, Marvin C. "Evaluation Theory Development." *Evaluation Comment* 2 (1969): 2-7.

American Association of University Professors. "1976 Recommended Institutional Regulations on Academic Freedom and Tenure." *AAUP Bulletin* 62 (Summer 1976): 184-91.

Anderson, Scarvia and Ball, Samuel. *The Profession and Practice of Program Evaluation*. San Francisco: Jossey-Bass, 1978.

Apple, Michael W. "The Process and Ideology of Valuing in Educational Settings." *Educational Evaluation: Analysis and Responsibility*, edited by M.W. Apple, M.J. Subkoviak, and H.S. Lufler. Berkeley: McCutchan, 1974.

Bailey, John J. and O'Connor, Robert J. "Operationalizing Incrementalism: Measuring the Muddles." *Public Administration Review* 35 (January/February 1975): 60-66.

Ball, Samuel. *Evaluating Educational Programs*. Princeton: Educational Testing Service, 1979.

Barak, Robert J. "A Survey of State-Level Academic Program Review Policies and Procedures for Higher Education. Final Report." Des Moines: Iowa State Board of Regents, 1975. ED 107 163. MF-$0.98; PC-$7.77.

_____ and Berdahl, Robert O. *State Level Academic Program Reviews in Higher Education*. Denver: Education Commission of the States, 1977. ED 138 638. MF-$0.98; PC-$10.77.

Berdahl, Robert O. "Evaluating Statewide Boards." *New Directions for Institutional Research*. Number 5. San Francisco: Jossey-Bass, 1975a.

_____. "Legislative Program Evaluation." *Increasing the Public Accountability of Higher Education*, edited by John K. Folger. San Francisco: Jossey-Bass, 1977, pp. 35-65.

Bogue, E.G. "Measuring Institutional Performance at the State Level." Paper presented at the annual conference of the American Association for Higher Education, Chicago: 1976. Mimeographed. HE 012 177 (June 1980 RIE). MF-$0.98; PC-$2.96.

——— and Troutt, W.E. "Allocation of State Funds on a Performing Criterion: Acting on the Possible While Awaiting Perfection." Paper presented at the annual meeting of the Association for Institutional Research, San Diego, 1977a. Mimeographed. HE 012 174 (June 1980 RIE). MF-$0.98; PC-$2.96.

——— and ———. "Bridging the Gap Between Accountability and Faculty Autonomy: A Campus-Based Approach in Developing Effectiveness Indicators." A paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978. Mimeographed. HE 012 172 (June 1980 RIE). MF-$0.98; PC-$4.46.

——— and ———. *Case Studies in Performing Funding*. Nashville: Tennessee Higher Education Commission, 1977b. HE 012 177 (June 1980 RIE). MF-$0.98; PC-$4.46.

Bloom, Benjamin S.; Hastings, J.T.; and Madaus George. *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill, 1971.

Braskamp, Larry A. "Evaluation in Higher Education: Departmental and Program Review. *Evaluation News* (Evaluation Network Newsletter) Number 12, Summer 1979, pp. 27-28.

———. "Student Satisfaction as a Useful Criterion in Departmental Review." Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978. Mimeographed.

Brzezinski, Evelyn J. "Quality in Reporting Evaluation Results: A Field Test of the Joint Committee's Draft Standards for Educational Evaluation." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979. Mimeographed.

Campbell, Donald T. and Cecil, Joe Shelby. "Protection of the Rights and Interests of Human Subjects in the Areas of Program Evaluation, Social Experimentation, Social Indicators, Survey Research, Secondary Analysis of Research Data, and Statistical Analysis of Data From Administrative Records. Evanston: Northwestern University, Department of Psychology, April 1977. Mimeographed.

——— and Stanley, J.C. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand-McNally, 1966.

Carnegie Council on Policy Studies in Higher Education. *Fair Practices in Higher Education*. San Francisco: Jossey-Bass, 1979.

Carroll, John B. "School Learning over the Long Haul." In *Learning and the Educational Process*, edited by I.D. Krumboltz. Chicago: Rand-McNally, 1965.

Caruthers, Kent and Orwig, Melvin. *Budgeting in Higher Education*. Washington, D.C.: American Association for Higher Education. ED 169 875. MF-$0.98; PC-$9.27.

Clark, Mary Jo. "Dimensions of Quality in Doctoral Education." Princeton: Educational Testing Service. *Findings* 1 (1974): 1-4.

———. "A Practical Guide to Graduate Program Review." Princeton: Educational Testing Service. *Findings* 5 (1979): 1-4. ED 155 959. MF-$0.98; PC-$2.96.

——. *Program Review Practices of University Departments.* Princeton: Educational Testing Service, 1977.

——; Hartnett, R.T.; and Baird, L.L. *Assessing Dimensions of Quality in Doctoral Education. A Technical Report of a National Study in Three Fields.* Princeton: Educational Testing Service, 1976. ED 173 144. MF-$0.98; PC-$30.02.

Clark, Woodrow W. "The Incorporating of Qualitative Techniques in Educational Evaluation." *CEDR Quarterly* 10 (1977): 8-10.

Clemmer, R.B. "Societal Indicators in Statewide Assessment." Salem, Oregon: Department of Education, 1977. Mimeographed.

Cohen, Audrey C. *The Citizen as the Integrating Agent: Productivity in the Human Services.* Rockville, Md.: Project Share, Monograph number 9, 1978.

Cohen, David and Garet, M. "Reforming Educational Policy With Applied Social Research." *Harvard Educational Review* 45 (1975): 17-43.

Comptroller General of the United States. *Problems and Needed Improvements in Evaluating Office of Education Programs.* Washington, D.C.: U.S. General Accounting Office, 1977.

Conrad, C. "University Goals: An Operative Approach." *Journal of Higher Education* 45 (1974): 505-515.

Cooley, William H. and Lohnes, Paul L. *Evaluation Research in Education.* New York: Halstead Press, 1976.

Cronbach, L.J. "Course Improvement Through Evaluation." *Teachers College Record* 64 (1963): 672-683.

——. "Remarks to the New Society." *Evaluation Research Society Newsletter* 1 (April 1977): 1, 2.

Cyert, Richard M. and March, James G. *A Behavioral Theory on the Firm.* Englewood Cliffs, N.J.: Prentice Hall, 1963.

De Neufville, Judith Innes. *Social Indicators and Public Policy.* New York: Elsevier Scientific Publishing, 1975.

Denny, T. "The Role of the Jury in the Judicial Evaluation Approach." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1976. Mimeographed. ED 151 403. MF-$0.98; PC-$2.96.

Deutscher, Irwin. "Toward Avoiding the Goal Trap in Evaluation Research." *Readings in Evaluation Research, Second Edition.* New York: Russell Sage Foundation, 1977.

Dougherty, Edward A. "The Role of Institutional Research in Program Discontinuance." Paper presented at the annual meeting of the Association for Institutional Research, San Diego, 1979a. Mimeographed. ED 174 081. MF-$0.98; PC-$2.96.

——. "What is the Most Effective Way to Handle Program Discontinuance? Case Studies From 10 Campuses." Paper presented at the annual meeting of the American Association of Higher Education, Chicago, 1979b. Mimeographed. HE 012 176 (June 1980 RIE). MF-$0.98; PC-$1.46.

Dressel, Paul and Simon, Lou Anna. *Allocating Resources Among Departments.* San Francisco: Jossey-Bass, 1976.

Dumont, Richard G. "Final Report for the FIPSE Performance Funding Project at Tennessee Technological University." Cookeville, Tennessee: Tennessee

Technological University, 1978a. HE 012 182 (June 1980 RIE). MF-$0.98; PC-$5.96.

———. Report on a Pilot Project in Performance Funding." Paper presented at the annual meeting of the Association for Institutional Research, Houston, 1978b. Mimeographed. ED 161 388. MF-$0.98; PC-$1.46.

Edwards, W.; Guttentag, M.; and Snapper, K. "A Decision-Theoretical Approach to Evaluation Research." In *Handbook of Evaluation Research*, edited by E.L. Struening and M. Guttentag. Beverly Hills: Sage Publications, 1975.

Fenno, Richard F., Jr. *Power of the Purse: Appropriations Politics in Congress*. Boston: Little, Brown, 1966.

Fincher, Cameron. *The Closing System of Academic Employment*. Atlanta: Southern Regional Education Board, 1978a. ED 160 032. MF-$0.98; PC-$7.77.

———. "Importance of Criteria for Institutional Goals." *Using Goals in Research and Planning*. San Francisco: Jossey-Bass, 1978b.

Fisher, G.H. "The Role of Cost-Utility Analysis in Program Budgeting." In *Program Budgeting*, edited by David Novick. Washington, D.C.: U.S. Government Printing Office, 1964.

Floden, Robert and Weiner, Stephen. "Rationality to Ritual: The Multiple Roles of Evaluation in Government Process." *Policy Sciences* 9 (February 1978): 9-18.

Florida Board of Regents. *Chance by Design*. Tallahassee: Florida Board of Regents, n.d.

Gardner, Don E. "Five Evaluation Frameworks: Implications for Decision Making in Higher Education." *Journal of Higher Education* 48 (1977): 571-593.

Gephart, William J. "The Coming Age of Evaluation." *CEDR Quarterly* 12 (Summer 1979): 2, 12, 13.

Glenny, Lyman; Bowen, Frank M.; Meisinger, Richard J., Jr.; Morgan, Anthony W.; Purves, Ralph A.; and Schmidtlein, Frank A. *State Budgeting For Higher Education: Data Digest*. Berkeley: University of California, Center for Research and Development in Higher Education, 1975. ED 139 303. MF-$0.98; PC-$27.02.

Gooler, Dennis D. "Evaluating Distance Education Programs." *Canadian Journal of University Continuing Education* 6 (Summer 1979): 43-55.

Gray, Robert G.; Jacobson, Edmund; Micek, Sidney S.; Patrick, Cathleen; Renkiewicz, Nancy; and Van Dusen, William. *Student Outcomes Questionnaires: An Implementation Handbook*. Boulder: National Center for Higher Education Management Systems, 1979. ED 172 671. MF-$0.98; PC-$15.58.

Gross, E. and Grambsch, P.V. *Changes in University Goals 1964-1971*. New York: McGraw Hill, 1974.

Guba, E.G. and Stufflebeam, D.L. "Evaluation: The Process of Stimulating, Aiding and Abetting Insightful Action." Paper presented at the Second Phi Delta Kappa National Symposium for Professors of Educational Research, Boulder, 1968. Mimeographed. ED 055 733. MF-$0.98; PC-$1.46.

Hall, Melvin. "Issues in Higher Education Program Evaluation." Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978. Mimeographed. ED 169 103. MF-$0.98; PC-$2.96.

Halstead, D. Kent. *Statewide Planning in Higher Education*. Washington, D.C.: Department of Health, Education, and Welfare, 1974. ED 096 914. MF-$1.92; PC-$34.65.

Hammond, R.L. "Evaluation at the Local Level. *Educational Evaluation: Theory and Practice*, edited by D. Stufflebeam et al. Worthington, Ohio: Charles Jones Publisher, 1971.

Harshman, Carl L. "A Model for Assessing the Quality of Nontraditional Programs in Higher Education." St. Louis: St. Louis University, 1979. Mimeographed. ED 175 321. MF-$0.98; PC-$10.46.

Hemphill, J.K. "The Relationship Between Research and Evaluation Studies." *Educational Evaluation: New Roles, New Means*. Chicago: National Society for the Study of Education, 1969.

Hill, Joan A.; Lutterbie, Patricia H.; and Stafford, Joseph. "Statewide Academic Program Review: The Florida Plan." Paper presented at the annual meeting of the American Association for Higher Education, Washington, D.C., 1979. Mimeographed. HE 012 180 (June 1980 RIE). MF-$0.98; PC-$2.96.

Hodgkinson, Harold L.; Hurst, Julie; and Levine, Howard. "Current Evaluation Practices in Innovative Colleges and Universities." *The Research Reporter* 9 (1975): 1-4. ED 104 263. MF-$0.98; PC-$2.96.

Hopkins, John E. "The JDRP: A Source of Concern." *Educational Research and Development Evaluators Newsletter* (American Educational Research Association) 5 (February 1, 1977): 3-6.

House, Ernest. "Assumptions Underlying Evaluation Models." *Educational Researcher* 7 (1978): 4-12.

Hutchinson, Thomas. "Some Sources of Error in the Efficiency of Evaluations." *CEDR Quarterly* 8 (1977): 15, 16, 23, 24.

Impara, James C. "Are Indicators the Yellow Brick Road?" Paper presented at the annual meeting of the American Educational Research Association, New York, 1977. Mimeographed.

Jackson, Gregg. "Adversary Evaluation—Sentenced to Death Without a Fair Trial." *Educational Researcher* 6 (1977): 2, 16.

Jaeger, Richard M. "An Abundance of Answers in Search of Questions: On a Methodology of Assessment Through Indicators." Paper presented at the annual meeting of the American Educational Research Association, New York, 1977. Mimeographed. ED 135 872. MF-$0.98; PC-$2.96.

———. "Measurement Consequences of Selected Standard Setting Models." Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, 1976. Mimeographed.

Jemelka, R.P. and Borich, Gary. "Traditional and Emerging Definitions of Educational Evaluation." *Evaluation Quarterly* 3 (1979): 263-276.

Jung, M.; Hamilton, J.A.; Helliwell, C.B.; and Wheeler, J.D. *A Study of State Oversight in Postsecondary Education*. Palo Alto: American Institutes for Research, 1977. ED 150 931. MF-$0.98; PC-$27.02.

Kourilsky, M. "An Adversary Model for Educational Evaluation." *Evaluation Comment* 4 (1973): 3-6.

Land, Kenneth C. and Spileman, Seymour (eds.). *Social Indicator Models*. New York: Russell Sage Foundation, 1975.

Lee, Eugene C. and Bowen, Frank M. *Managing Multicampus Systems*. San Francisco: Jossey-Bass, 1975.

——— and ———. *The Multicampus University: A Study of Academic Governance*. Highstown, N.J.: McGraw-Hill, 1971.

Lenning, Oscar. *Previous Attempts to Structure Educational Outcomes and Out-*

come-Related Concepts: A Compilation and Review of the Literature. Boulder: National Center for Higher Education Management Systems, 1977.

———; Lee, Yong S.; Micek, Sidney S.; and Service, Allen L. *A Structure for the Outcomes of Post-secondary Education*. Boulder: National Center for Higher Education Management Systems, 1977. ED 150 904. MF-$0.98; PC-$7.77.

Lent, Richard. "Planning for the External Evaluation of the University of Mid-America." Lincoln, Nebraska: University of Mid America, Office of Research and Evaluation, 1974. Mimeographed. ED 159 963. MF-$0.98; PC-$4.46.

Levin, H.M. "Cost-Effectiveness in Evaluation Research." *Handbook of Evaluation Research*, edited by E.L. Struening and M. Guttentag. Beverly Hills: Sage Publications, 1975.

Levine, M. "Scientific Method and the Adversary Model: Some Preliminary Suggestions." *American Psychologist* 31 (1974): 661-677.

Lifson, M.W. "Value Theory." *Cost Effectiveness: The Economic Evaluation of Engineered Systems*. New York: John Wiley & Sons, 1968: 79-112.

Lindvall, C.M. "Basic Considerations in Assessing the Validity of Evaluation Designs." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979. Mimeographed. ED 170 369. MF-$0.98; PC-2.96.

Lingenfelter, Paul E. "The Politics of Higher Education Appropriations in Three Midwestern States." Unpublished dissertation. Ann Arbor: University of Michigan, 1974.

March, James G. and Simon, Herbert A. *Organizations*. New York: John Wiley & Sons, 1958.

McCartan, Anne-Marie. "Future May Be Now in California." *Community and Junior College Journal* 49 (1979): 36-40.

Meeth, L. Richard. *The Impact of State and Federal Funding Regulations on Nontraditional Postsecondary Education*. Washington, D.C.: George Washington University, Institute for Educational Leadership, 1975.

Meisinger, Richard J., Jr. *State Budgeting for Higher Education: the Use of Formulas*. Berkeley: University of California, Center for Research and Development of Higher Education, 1976. ED 132 963. MF-$0.98; PC-$20.40.

Michigan State Department of Education. "Formula Funding Mechanisms for State Support of Public Colleges and Universities in Michigan Based on a Study of Funding Mechanisms Across the Nation." Lansing: Michigan State Department of Education, 1976. ED 131 711. MF-$0.98; PC-$5.96.

Mims, R. Sue. "Program Review and Evaluation: Designing and Implementing the Review Process." Paper presented at the annual meeting of the Association for Institutional Research, Houston, 1978. Mimeographed.

Mingle, J.R. "Influencing Academic Outcomes: The Power and Impact of Statewide Program Reviews." In *The Closing System of Academic Employment*, edited by Cameron Fincher. Atlanta: Southern Regional Educational Board, 1978. ED 160 032. MF-$0.98; PC-$7.77.

Minmier, George and Hermansen, Robert H. "A Look at Zero-Base Budgeting —The Georgia Experience." *Atlanta Economic Review* 26 (1976): 5-12.

Morris, Lynn L. and Fitz-Gibbon, Carol T. *How To Present an Evaluation Report*. Beverly Hills: Sage Publications, 1978.

Moss, Charles E. and Gaither, Gerald H. "Formula Budgeting: Requiem or Renaissance?" *Journal of Higher Education* 47 (1976): 513-564.

Murphy, Jerome T. "Legislative Program Review: A Guide for Legislative Committee Staff." New Brunswick: Rutgers University, Eagleton Institute, 1976. Mimeographed.

Orlans, Harold. *Private Accreditation and Public Eligibility*. Lexington, Mass.: Lexington Books, 1975.

Owens, T.R. "Educational Evaluation by Adversary Proceeding." In *School Evaluation: The Politics and the Process*. Berkeley: McCutchan Publishers, 1973.

_____ and Hiscox, M.D. "Alternative Models for Adversary Evaluation: Variations on a Theme." Paper presented at the annual meeting of the American Educational Research Association, 1977. Mimeographed. ED 136 425. MF-$0.98; PC-$4.46.

Parlett, Malcolm and Hamilton, David. "Evaluation as Illumination: A New Approach to the Study of Innovatory Programs." In *Evaluation Studies Review Annual, Volume 1*, edited by Gene Glass. Beverly Hills: Sage Publications, 1976.

Patton, Michael Q. *Utilization-Focused Evaluation*. Beverly Hills: Sage Publications, 1978.

Peterson, Marvin W. "The State Level Performance Assessment Process: Concepts, Perspectives and Issues." Ann Arbor: University of Michigan, Center for Higher Education, 1977. Mimeographed.

Peterson, R.E. and Uhl, N.P. *Formulating College and University Goals: A Guide for Using the IGI*. Princeton: Educational Testing Service, 1977.

Pethel, R. and Brown, R. "A Matter of Facts: State Legislative Performance Auditing." *Public Administration Review* 31 (July/August 1974): 318-327.

Popham, James. *Educational Evaluation*. Englewood Cliffs, N.J.: Prentice Hall, 1975.

_____ and Carlson, D. "Deep Dark Deficits of the Adversary Evaluation Model." *Educational Researcher* 6 (1977): 3-6.

Poulton, Nick L. "Program Review and Evaluation: Integrating Results Into Decision Making." Paper presented at the annual meeting of the Association for Institutional Research, Houston, 1978. Mimeographed. HE 012 178. (June 1980 RIE). MF-$0.98; PC-$2.96.

Provus, Malcolm. *Discrepancy Evaluation*. Berkeley: McCutchan, 1971.

Pyhrr, Peter A. *Zero-Base Budgeting*. New York: John Wiley & Sons, 1973.

Ramsey, James R. and Hackbart, Merlin M. "Budgeting: Inducements and Impediments to Innovations." *State Government* 46 (Spring 1979): 65-69.

Rippey, R.M. (ed.). *Studies in Transactional Evaluation*. Berkeley: McCutchan, 1973.

Romney, Leonard C. and Micek, Sidney. "Translating Goals Into Measurable Objectives: Research Studies and Practical Procedures." Paper presented at the annual meeting of the Association for Institutional Research, Montreal, 1977. Mimeographed. ED 140 688. MF-$0.98; PC-$2.96.

Rose, Clare and Nyre, Glen. "Holistic Evaluation" An Eclectic Approach to Program Evaluation. Paper presented at the annual meeting of the American Educational Research Association, New York, 1977. Mimeographed.

Rose, Lee and Cronbach, L. J. "Handbook of Evaluation Research." *Educational Researcher* 5 (1976): 9-19.

Rossi, Robert and Gilmartin, Kevin. "Social Indicators and Social Forecasting." *Social Indicators* Number 12 (1979): 1-10.

Sanders, James R. "On the Technology and Art of Evaluation: A Review of Seven Evaluation Primers." *Evaluation News* (Evaluation Network Newsletter) Number 12 (Summer 1979): 11-16.

Schick, Allen. *Budget Innovation in the States*. Washington, D.C.: The Brookings Institution, 1971.

———. "Zero-Base Budgeting and Sunset: Redundancy or Symbiosis?" *The Bureaucrat* 6 (1977): 12-32.

Scriven, Michael. "Evaluation Bias and Its Control." *Evaluation Studies Review Annual, Volume 1*, edited by Gene Glass. Beverly Hills: Sage Publications, 1976.

———. "The Methodology of Evaluation." In *Curriculum Evaluation*, edited by Ralph Tyler. Chicago: Rand McNally, 1967.

———. "Prose and Cons About Goal Free Evaluation." *Evaluation Comment* 3 (December 1972): 1-4.

Sigelman, Lee. "The Quality of Administration: An Exploration in the American States." *Administration & Society* 8 (May 1976): 107-144.

Sharkansky, Ira. "Agency Requests, Gubernatorial Support and Budget Success in State Legislatures." *American Political Science Review* 62 (1968): 1221-1231.

Sheiring, Michael J. "Zero-Base Budgeting in New Jersey." *State Government* 49 (Summer 1976): 174-179.

Sherman, Betty. *Making Government Work: A Common Cause Report on State Sunset Activity*. Washington, D.C.: Common Cause, 1979.

Shirley, Robert C. and Volkwein, J. Fredericks. "Establishing Academic Program Priorities." *Journal of Higher Education* 49 (September/October 1978): 472-488.

Smith, Nick and Murray, Stephen L. *Techniques and Procedures for Formative Evaluation*. Portland, Oregon: Northwest Regional Educational Laboratory, 1975.

Smoke, H. Richard. "The Practice of Evaluation as a Purposeful Activity." Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., 1975. Mimeographed. ED 102 700. MF-$0.98; PC-$2.06.

——— and Hake, Harold. "Cope: A Systematic Approach to the Evaluation of Academic Departments." Paper presented at the annual meeting of the American Educational Research Association, New York, 1977. Mimeographed. ED 146 180. MF-$0.98; PC-$2.06.

Sprenger, J.M. and Scultz, R.E. "Staff Reduction Policies." *College Management* 9 (May 1974): 22, 23.

Stake, R.E. "The Countenance of Educational Evaluation." *Teachers College Record* 68 (1967): 523-540.

———. *Program Evaluation, Particularly Responsive Evaluation*. Kalamazoo, Mich.: Western Michigan University, The Evaluation Center, 1975.

——— and Denny, T. "Needed Concepts and Techniques for Utilizing More Fully the Potential of Evaluation." In *Educational Evaluation: New Roles,*

New Means, edited by Ralph Tyler. Chicago: National Society for the Study of Education, 1969.

——— and Gjerde, C. *An Evaluation of TCITY: the Twin City Institute for Talented Youth*. Kalamazoo, Michigan: Western Michigan University, The Evaluation Center, 1975.

Steinbruner, John D. *The Cybernetic Theory of Decision: New Diemnsions of Political Analysis*. Princeton: Princeton University Press, 1974.

Struenign, E.L. and Guttentag, M. (eds.). *Handbook of Evaluation Research*. Beverly Hills: Sage Publications, 1975.

Stufflebeam, Daniel; Foley, W.J.; Gephart, W.J.; Guba, E.G.; Hammond, R.L.; Merriman, H.O.; and Provus, M.M. *Educational Evaluation and Decision Making*. Itasca, Ill.: Peacock Publishers, 1971.

Suchman, E. *Evaluation Research*. New York: Russell Sage Foundation, 1972.

Tallmadge, G. Kasten. *The Joint Dissemination Review Panel Ideabook*. Mountain View, California: RMC Research Corporation, 1977. Available from U.S. Office of Education. ED 148 329. MF-$0.98; PC-$9.27.

Task Force on Coordination, Governance and Structure of Postsecondary Education. *Coordination or Chaos?* Denver: Education Commission of the States, 1973. ED 093 239. MF-$0.98; PC-$9.27.

Tennessee Higher Education Commission. *The Competent College Student*. Nashville: THEC, 1977.

Terenzini, Patrick T.; Hartmark, Leif; Lorang, Wendell; and Shirley, Robert C. "A Conceptual and Methodological Approach to the Identification of Peer Institutions." Paper presented at the annual meeting of the Association for the Study of Higher Education, Washington, D.C., 1979.

Thorndike, R.L. and Hagen, E. *Measurement and Evaluation in Psychology and Education*. New York: John Wiley & Sons, 1961.

Thurston, P. "Revitalizing Adversary Evaluation: Deep Dark Deficits or Muddled Mistaken Musings." *Educational Researcher* 7 (1978): 3-8.

Tyler, Ralph. *Basic Principles of Curriculum and Instruction*. Chicago: The University of Chicago Press, 1950.

Troutt, William E. "Regional Accreditation Evaluative Criteria and Quality Assurance." *Journal of Higher Education* 50 (1979): 199-210.

Van Alstyne, Carol. "Postsecondary Education Indicators: Can We Measure Our Value Imperatives—Access, Diversity, and Excellence?" In *Planning, Managing and Financing in the 1980's*. Boulder: National Center for Higher Education Management Systems, 1978. ED 165 610. MF-$0.98; PC-$10.77.

Weiss, Carol H. "Between the Cup and the Lip." *Evaluation Comment* 1 (1973): 1-6.

———. "Evaluating Educational and Social Action Programs: A Tree Full of Owls." In *Evaluating Action Programs*, edited by Carol Weiss. Boston: Allyn & Bacon, 1972: 3-27.

Wildavsky, Aaron. *The Politics of the Budgetary Process*. Boston: Little, Brown, 1964.

Windle, Charles. "Resources for Program Evaluation: Cultivating Skills for Independence (and Dependence)." Paper presented at the American Psychological Association convention, Toronto, 1978. Mimeographed.

Wolf, R.L. "Evidence: Educational Evaluation and the Metaphors of Law." Paper presented at the annual meeting of the American Educational Research Association, Washington, 1975a. ED 115 673. MF-$0.98; PC-$2.96.

Wolf, R.L. "Trial by Jury: A New Evaluation Method." *Phi Delta Kappan* 57 (1975b): 185-187.

Worthen, Blaine R. and Sanders, James R. *Educational Evaluation: Theory and Practice*. Worthington, Ohio: Charles A. Jones Publishing, 1973.

Wood, Lynn and Davis, Barbara G. *Designing and Evaluating Higher Education Curricula*. Washington, D.C.: American Association for Higher Education, 1978. ED 165 669. MF-$0.98; PC- $5.96.

Woy, J. Richard; Lund, Donald A.; and Attkisson, C. Clifford. "Quality Assurance in Human Services Program Evaluation." In *Evaluation of Human Service Programs*, edited by C. Clifford Attkisson. New York: Academic Press, 1978.

Reports Published in the 1979 Series

1. Women in Academe: Steps to Greater Equality
   Judith Gappa and Barbara Uehling

2. Old Expectations, New Realities: The Academic Profession Revisited
   Carol Herrnstadt Shulman

3. Budgeting in Higher Education
   J. Kent Caruthers and Melvin Orwig

4. The Three "R's" of the Eighties: Reduction, Retrenchment, and Reallocation
   Kenneth P. Mortimer and Michael L. Tierney

5. Occupational Programs in Four-Year Colleges: Trends and Issues
   Dale F. Campbell and Andrew S. Korim

6. Evaluation and Development of Administrators
   Robert C. Nordvall

7. Academic Advising: Getting Us Through the Eighties
   Thomas J. Grites

8. Professional Evaluation in the Eighties: Challenges and Response
   Glenn F. Nyre and Kathryn C. Reilly

9. Adult Baccalaureate Programs
   Marilou Denbo Eldred and Catherine Marienau

10. Survival Through Interdependence: Assessing the Cost and Benefits of Interinstitutional Cooperation
    Lewis D. Patterson

Reports Published in the 1980 Series

1. Federal Influence on Higher Education Curricula
   William V. Mayville

To subscribe to the Research Report series (10 issues starting from date of subscription), write to the Publications Department, American Association for Higher Education, One Dupont Circle, Suite 780, Washington, D.C. 20036. The subscription rate for AAHE members is $25, for nonmembers $35. Payment must accompany all orders under $15. Single copies of 1979 and 1980 Research Reports—$3 for AAHE members, $4 for nonmembers. Bulk discounts available.