Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Feature Article—Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis

Saul I. Gass,

Feature Article

Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis

SAUL I. GASS

University of Maryland, College Park, Maryland

(Received June 1981; revised October 1982; accepted November 1982)

The extension of OR decision-aiding models and OR methodology from operational and technological settings to the field of policy analysis has caused analysts and users to question whether the OR process can handle the requirements of this new area. This paper reviews the evolving nature of OR methodology, the difficulties that have arisen in applying it, and the attempts to improve the use of OR methodology in policy analysis. The analyst's role in providing information to the user to determine whether and how a specific OR model can be used as a decision aid is emphasized. Relevant information for the analyst to provide covers model validation, assessment, utility, confidence and documentation.

OPERATIONS RESEARCH (OR) is a scientific approach to decision-making involving the operations of organizational systems (Hillier and Lieberman [1974]); or more succinctly, operations research is the science of decision-making. *The scientific contribution* of OR is in the development of decision-aiding models.

The OR resolution of a decision problem is based on the analyst's ability to translate the decision problem into a form—usually a mathematical model—that can be used to compare the extent to which each alternative solution satisfies the problem objective. How to determine the adequacy of a given model to aid in such a comparison is a major concern of this paper.

1. THE MODEL DEVELOPMENT PROCESS

An early formalization of the OR model and methodology was given by Goode [1957]. See Figure 1. The 20 years from 1950 to 1970 were most productive in utilizing "standard" OR methodology. New decision problems were solved in all areas of business, industry, and government. Advanced computational procedures, coupled with the expanded power of the computer, increased our abilities. The improvements continue

Subject classification: 91 decision analysis, 251 government, 531 philosophy of modeling, 608 OR/MS standards.

Operations Research
Vol. 31, No. 4, July–August 1983

0030-364X/83/3104-0603 $01.25

today. Those 20 years were most successful as the mainstream of OR analysts concentrated on solving operational decision problems, the *raison d'être* of the OR methodology. (See survey by Ledbetter and Cox [1977].)

Policy Analysis and "Squishy" Problems

Beginning in the late 1960s OR-based methodology was applied to a different class of decision problems, that is, problems involving public policy and business policy issues. "Policy analysis" refers to the systematic analysis of questions faced by the governmental planning or

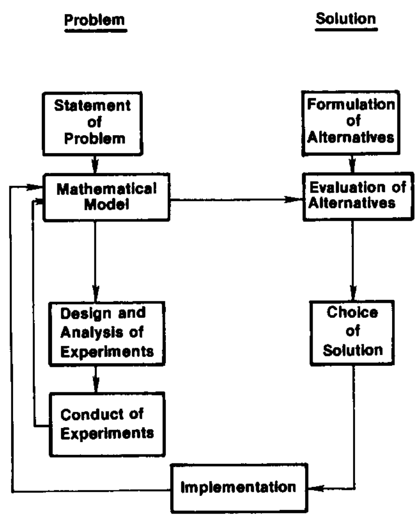

Figure 1. Role of the mathematical model in operations research methodology. (From Goode [1957]; the implementation box has been added.)

decision-making process, conducted with the intention of affecting that process (Strauch [1974]). Policy analysts tend to organize their investigations around a set of policy objectives. They investigate which program or combination of programs might achieve the desired ends most efficiently (Greenberger et al. [1976]). The outcome of a policy analysis is rarely a clear recommendation for choice, Quade [1982]. As noted by Strauch, "... problems encountered in policy analysis run the gamut from well-defined to highly 'squishy.' A well-defined problem is one that can be given a clear-cut, well-defined formulation, amenable to rigorous analysis. A squishy problem is one with the property that any clear-cut, well-defined formulation of it will look like an unambiguous representation of the substantive problem only so long as we don't lean too hard on it, or question it too carefully or deeply ... well-defined problems in policy analysis tend to be concerned almost exclusively with technological or physical questions. Problems with any significant degree of behavioral or political content tend to be squishy. The more central the behavioral or political content becomes ..., the squishier the problem is likely to be."

A favorite tool for investigating policy problems has been the OR methodology and its set of mathematical decision models. The ability of OR (more generally, systems analysis) to resolve policy analysis decision problems, both in and out of government, is highly questionable (Hoos [1972], Brewer [1973], Brewer and Shubik [1979], Fromm et al. [1975], U.S. General Accounting Office (U.S. GAO) [1971], Orden [1979]). In contrast to OR's success in the operational problem area (problems that tend to be well-defined and "hard"), the modeling of policy analysis decision problems has not lived up to expectations. There are many reasons for this (see discussion in Greenberger et al.). The reason we consider most important is the inability of the analyst to demonstrate to potential users and critics that a model and its results have any credibility within the policy decision environment. It is the lack of credibility that makes the difference.

For operational problems, the results of an OR model-based analysis usually can be accepted or rejected by submitting the new decision information to a face validity test (Hermann [1967], Emshoff and Sisson [1970]). This test is the surface or initial impression of a model's realism and is obtained by asking people who know the real system (managers, production engineers, etc.) to judge whether the model and its results are reasonable. The key point here is that the recommended solution can be tested in a working environment. For example, feed mix can be felt, smelled, and fed to cattle; another cashier's window can be opened in the bank; the oil refinery can operate effectively under the recommended actions; or the inventory system can be used to satisfy customer demands.

In contrast, there are important policy analysis problems, whose modeling resolutions defy a face validity test. These include economic analyses attempting to describe market behavior, war-gaming, urban planning, analyses on a global scale, and most problems with a human behavioral component. The OR decision model has been used to analyze many such policy problems. The lack of credibility of most of the related models contrasts sharply with the wholesale acceptance and use of the OR model in operational problem areas. There is an inverse relationship between a model's credibility and the squishiness of the problem. The more squishy, the less credible. This does not mean that OR methodology should not be used as a decision aid in the resolution of policy analysis problems. It does mean that OR analysts must work at ways to produce evidence of credibility for policy models. This has been done for some policy areas, but the applications tend to be operational and technological (Larson and Odoni [1981], Walker et al. [1979], Quade [1966, 1982]).

A Current View of the OR Process

To maintain the success of OR as the science of decision-making calls for continuous reappraisal of OR methodology. The problems have become larger and more complex, and the broadening of OR to include policy analysis problems highlights the need for credibility. Data issues are most important. The model is no longer a paper mathematical statement or a mental model, but is the programmed computer-based model. Model validation and assessment (evaluation) are important aspects of the analysis and the use of the results. A current view of the modeling process encompassing most of these aspects is given in Figure 2 (U.S. GAO [1979]). (We emphasize that the steps of this process are not sequential, many are parallel, and documentation of the model begins at the beginning of the process.)

An Extended View of the OR Process

For models used to analyze public policy issues, the modeling process must be extended further, as shown in Figure 3 (Clark and Cole [1975]). It is this approach to model building that is our concern. The basic modeling steps, shown here in an aggregated fashion stressing verification and validation, are augmented by public debate, evaluation by policy advisers, and independent model assessment. Past attempts to accomplish these steps for policy studies of ABM deployment, energy, urban development, limits of growth, caused issues of validation and assessment to be raised that questioned the value of the total modeling process and OR methodology for public policy analyses. This motivated federal researchers, in particular, to improve the decision-aiding uses of policy models. (See U.S. GAO [1976a, 1977, 1979], Professional Audit Review Team (PART) [1977, 1979], Shubik and Brewer [1972], Brewer and Shubik, Fromm et al., Greenberger et al., Gass and Sisson [1975], Gass [1980, 1981a], Brewer [1973], Hoos, Lee [1973] and Peck [1974].)

This paper was written to indicate how the research community concerned with policy models is attempting to develop and test procedures for improving the role of models as decision aids. We do not mean to imply that policy models will rest on as sound a footing as operational models. Nonetheless there is potential for improvement, and efforts in that direction are worthwhile.

Central to the proper development of any computer-based model is how the interrelated steps illustrated in Figures 2 and 3 are accomplished. The completeness and effectiveness of the model depends on the planned use of the model results, the available resources and time, and the inclinations and experiences of project sponsors, analysts, and programmers. This paper focuses on improving model development for policy

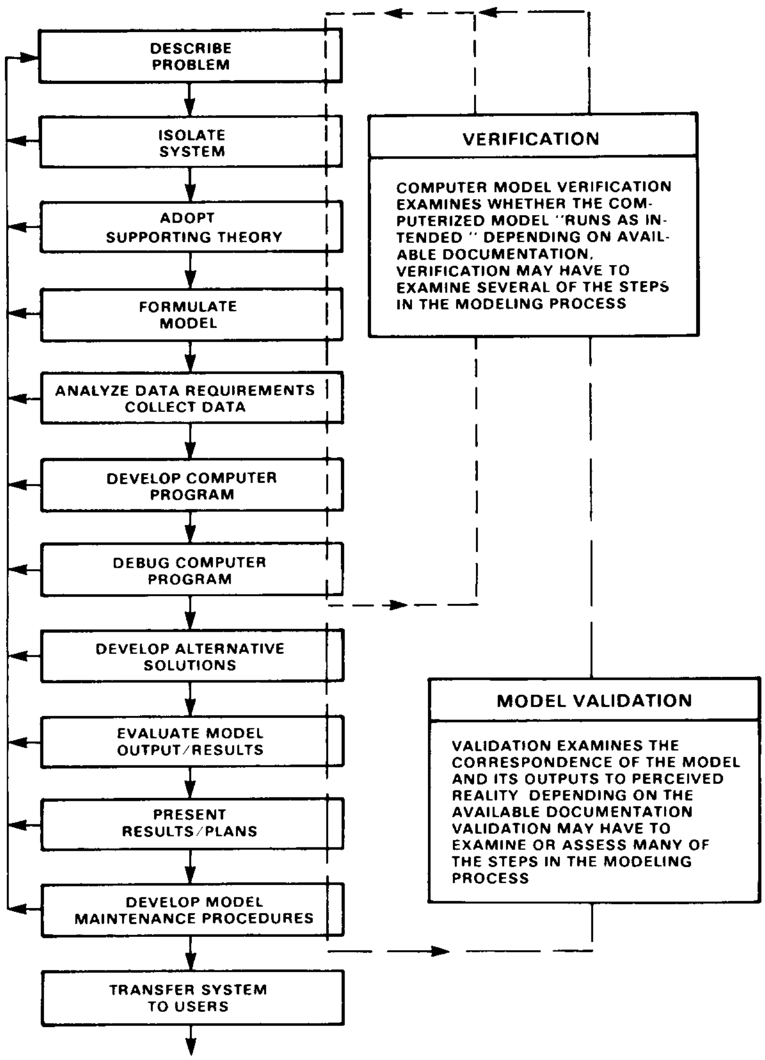

Figure 2. Basic steps in the modeling process. (From U.S. GAO [1979].)

analysis in terms of model validation, assessment and utility. (Although of equal importance, we do not address the creative aspects of "squishy" problem analysis and the model management procedures that can also advance the development process.)

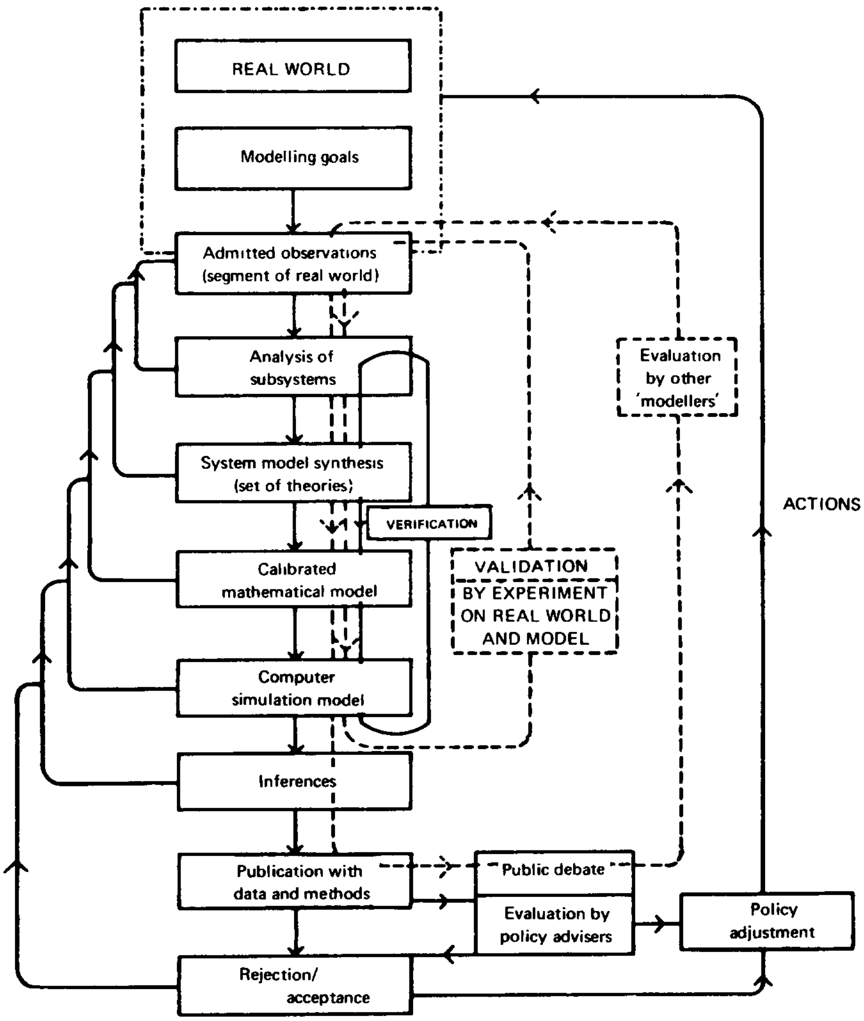

Figure 3. The model building process and its interaction with "the real world." (Modified with permission from Clark and Cole [1975]: *Global Simulation Models*, John Wiley & Sons, New York; the headings of the verification and validation boxes have been interchanged to agree with the definitions used in this paper.)

Verification

For computer-based models, the data inputs, the computer programs executed for a specified policy analysis, and the documentation comprise the model. (More correctly, the total modeling milieu [personnel, resources, theory, assumptions, algorithms, code, etc.] is the model.) There must be explicit procedures that measure the relationship between verbal and mathematical model descriptions, that often change over time, with the "final" programmed representation. The process of demonstrating that the computer program "runs as intended" is termed "verification." Verification involves (1) ensuring that the program, as written, accurately describes the model as designed, (2) ensuring that the program is properly mechanized on the computer, and (3) ensuring that the program as mechanized runs as intended (Fishman and Kiviat [1967], House and McLeod [1977]).

Verification tests fall into two indistinct categories: experiments to debug the logic of the computer program, and checks to demonstrate the correctness of the numerical and data procedures as carried out by the computer program. There is general agreement that for large and complex computer programs (and most models of public systems are of this type) it is impossible to demonstrate that a program has been completely verified. Verification tends to be the domain of the programmers, with some assistance from the analysts in devising numerical tests. (Some computer scientists advocate replacing verification with the concept of program reliability, that is, the program works for all practical purposes even though it is not perfect; see DeMillo et al. [1979]. In discussing complex computer programs used to design nuclear reactors, Griffin [1970] notes that "... it is apparent there is no such a thing as absolute verification of a computer program... Rather than talking about verification, it would seem more appropriate to talk about *level of confidence*." His program confidence is based on (a) qualifications of the programmer, (b) mathematical model and solution method, (c) documented verification, (d) length of time program has been in use, (e) diversity of use, and (f) current use. In a related paper, Griffin [1971] discusses the process of qualification (validation of the total analysis) of a solution method for solving a particular problem by a particular user. See our discussions on model validation and confidence that follows. The term software engineering is used by computer scientists to describe the process of producing reliable programs (Zelkowitz [1978].)

Validation

Model "validation" tests the agreement between the behavior of the model and the real world system being modeled (Fishman and Kiviat). For a decision-aiding model, validation tends to be the overriding concern of the analyst. How to measure the validity of a model is an open question. The answer depends on the real world aspect being analyzed, the type of model being used, who is asking the question, and who will interpret the answer. Orden notes that linear programming models of engineering and physical systems (for example, oil refineries and transportation networks) have physical validity, that is, results can be shown to work within the real world environment. Similarly, a feed-mix can be fed to chickens, or a proposed hotdog formula can be tasted. A simulation model of an ongoing physical system must be able to replicate the past to be acceptable as a decision aid (Emshoff and Sisson). But what of simulation and policy analysis models of nonexistent systems or for future-oriented situations (such as models for determining investments in energy technology) in which the past is not a good predictor of the future? For first-time or futuristic models validity is superceded by the concept of model credibility, as defined by the decision maker (Hermann, and Emshoff and Sisson). We shall return to this point below. As with program verification, analysts do not believe that a model can be completely validated (House and McLeod, House and Ball [1980]). Quade [1980] states: "A particularly dangerous myth is the belief that a policy model can be fully validated—that is, proved correct. Such models can, at best, be invalidated." And, "Thus the aim of the validation (or rather invalidation) attempts is to increase the degree of confidence that the events inferred from the model will, in fact, occur under the conditions assumed. When you have tried all the reasonable invalidation procedures you can think of, you will not, of course, have a valid model (and you may not have a model at all). You will, however, have a good understanding of the strengths and weaknesses of the model, and you will be able to meet criticisms of omissions by being able to say why something was left out and what difference including it would have made. Knowing the limits of the model's predictive capabilities will enable you to express proper confidence in the results obtained from it." Mindful of the preceding caveat, we shall courageously use the phrase "to validate a model" and discuss validation procedures below.

Assessment/Evaluation

Increasing use of policy models as aids in resolving controversial problems has encouraged a parallel effort in developing and applying methodologies for model "assessment" or "evaluation." (We use these terms interchangeably.) Greenberger and Richels [1979] define model assessment to include (1) verification, validation, and quality control of the usability of the model and its readiness for use and (2) investigations into the assumptions and limitations of the model, its appropriate uses, and why it produces the results it does (see also Gass [1977a, b], U.S. GAO [1979], Gass and Thompson [1980]). The basic idea behind any model assessment methodology is the accumulation evidence regarding the credibility and applicability of the model by an independent, interested party.

Utility

Models are developed for many purposes—research, experimentation, education, and decision-making. For all of these purposes the following definition of model utility or usability applies (U.S. Library of Congress [1975]). The "utility" of a model depends on its being useful and usable, and that it is used. A model can be considered *useful* if it shows promise of attaining the objectives initially established for it. A model can be considered *usable* if it is understandable and plausible to both technicians and policymakers, economic to solve in terms of data collection and computation, and accessible to those who wish to use it. Finally, if a model is useful and usable, it stands a good chance of being *used*, especially if potential users receive the proper training, supplemented by adequate documentation. For decision-aiding models, a model is useful only when the decision maker believes (rightly or wrongly) that it is valid (Emshoff and Sisson). We review next in more detail the concepts of validity, assessment, and related topics.

2. MODEL VALIDATION

Simply put, validation of a model attempts to establish how closely the model mirrors the perceived reality of the model user/developer team. The definition and the determination of the "level" of validation is the major task of model assessment.

As noted earlier, a first-time model, a model of a nonexisting system, or one that makes assumptions about the possible states of the future, is most difficult to validate. The following concepts apply:

1. Face validity or expert opinion: Is the initial impression of the model's realism positive when reviewed by decision makers who know the system being modeled? Is the model credible?

2. Variable-parameter validity or sensitivity analysis: How do the model's variables and parameters compare with their assumed counterparts in the real world? As data are changed, how are outputs affected?

3. Hypothesis validity: Do pairwise or higher level relationships correspond to similar relationships in the real world? Do the subsystem models interact in a realistic fashion? (See Herman, and Emshoff and Sisson.)

These concepts are, of course, applicable to all models.

Models of real systems can be subjected to additional tests. A real system is some part of the real world which is of interest; a real system produces behavioral data over time; a model is a set of (computer) instructions for generating behavioral data. Validity is measured by how well the real-system data compares with the model-generated data. The model is *replicatively valid* if it matches data already acquired from the real system. The model is *predictively valid* when it can match data before the data are acquired from the real system. A model is *structurally valid* if it not only reproduces the observed real system behavior, but truly reflects the way in which the real system operates to produce this behavior (Ziegler [1976]). Simulation texts such as Ziegler [1976], Shannon [1975], Law and Kelton [1981] and others describe standard statistical procedures for hypothesis testing and estimation that can be used for validating a model. These procedures include tests of means, analysis of variance or covariance, goodness of fit tests, regression and correlation analysis, spectral analysis, confidence intervals. However, the process of validating a model, especially a policy model, must go beyond applicable statistical tests. Thus we extend the above into a more general validation framework and discuss the integrated concepts of technical validity, operational validity, and dynamic validity according to Schellenberger [1974].

Technical Validity

Technical validity requires the identification of all model assumptions, including those dealing with data requirements and sources, and their divergences from perceived reality. Technical validity is a summation of results in terms of model validity, data validity, logical and mathematical validity, and predictive validity.

Model validity refers to the correspondence of the model to the real world and is concerned with identifying all stated and implied assumptions, identification and inclusion of all decision variables, and hypothesized relationships between variables. Assumptions can be grouped as (1) mathematical assumptions including the model form and continuity of the relationships, (2) content assumptions dealing with the scope and definition of model terms and variables, and (3) causal assumptions concerning assumed or hypothesized relationships between terms and variables. The analyst compares each assumption and hypothesis to the internal and external problem environments as viewed by the decision maker, and makes some statement regarding the extent of divergence.

Data validity deals with raw data and structured data. (Structured data are raw data upon which some manipulation has been performed, for example, aggregation.) Raw data validity is concerned with problems of measurement and determining if the data are true in terms of (1) accuracy (the ability to correctly identify, obtain, and measure what is desired), (2) impartiality (the assurance that the data are recorded correctly), and (3) representativeness (the assurance that the universe from which any sample data are drawn is properly identified and that the sample was random). Structured data validity requires review of each step of the manipulation and is a part of model verification. See Altman [1980] on the pitfalls of data analysis and Hirshfeld [1980] on data auditing.

Logical/mathematical validity is concerned with translating the model form into a numerical, computer process that produces solutions. This involves aspects of model verification such as (1) determining if the mathematical and numerical calculations are correct and accurate, (2) analyzing if the logical flow of data and intermediate results are correct, and (3) ensuring that variables and relationships have not been omitted. There is no standard methodology for determining logical validity. Approaches include comparison of model outcomes with expected or historical results, and a close examination of the model form and its numerical representation on a flow-chart.

The final element of technical validation is *predictive validity*, that is, analysis of errors between actual outcomes and predicted outcomes for a model's components and relationships. Outcomes can be parameters used as part of the internal computational process of the model (a demand forecast based on a time-series analysis), or an alternative solution selected by the model's evaluative process (the predicted profits based on a production schedule). Here we are looking for errors and their magnitudes, the reasons they exist, and if and how they can be corrected. This can be accomplished by statistical tests, comparisons with historical data, predictive accuracy over time, etc.

Thus far we have focused on the determination of the model's overall technical validity as a function of four components: model validity, data validity, logical validity, and predictive validity. Tests applied to each component produce information, the sum of which should enable the analyst (and others) to evaluate the technical validity of the model. The analyst will probably not make a definitive judgment whether or not the model is "valid," but the validity analysis will suggest areas of concern regarding the model's technical validity. In general, the appraisal should apply to all concerns regarding technical, operational, and dynamic validity.

Operational Validity

Operational validity attempts to assess the importance of the errors found under technical validity. Because models are inherently unable to totally reproduce or predict the real environment, operational validity must conclude whether the use of the model is appropriate for the observed and expected errors. The analysis should produce information that will enable the decision maker to conclude whether to accept or reject the model's solution. That is difficult to do. It requires close interaction between the developers and the decision maker to ensure that basic assumptions and trial outcomes are consistent with the decision maker's expectations.

The analyst must also consider the following elements of operational validity. Does the proposed new system offer a reasonable improvement in terms of net cost savings? We must recognize that new system costs are difficult to estimate and that small savings should be given little value. Many federal policy models compare existing systems with proposed systems by means of a benefit-cost criterion. The analyst must demonstrate that costs are attributed correctly, the method for determining the benefits is valid, and the stated benefits are attainable.

Operational validity is also concerned with whether the model can produce unacceptable answers for proper ranges of parameter values. That is, the model should be robust enough that a user would find it difficult to make the model yield an ostensibly wrong solution for the stated ranges of parameters.

Sensitivity analysis is another element of operational validity. In conducting a sensitivity analysis, we systematically vary the values of the model parameters over some range of interest to determine if and how the recommended solution changes (Shannon). This is done by running the model with extreme values of the parameters (usually varying one parameter at a time) and comparing the (possibly new) solution set of decision variables and their values with the recommended solution. If the solution turns out to be sensitive to certain parameter changes, the decision maker would probably want to require the model analysts to explore further or to justify in detail values of these parameters. If we are dealing with an optimizing model, we might find that the optimizing criterion is not sensitive to relatively large changes in the value of a decision variable, thus the need for a finely tuned set of parameters might be alleviated. The decision maker can then act with greater assurance that the implementation of any solution close to the recommended one will yield acceptable results.

Sensitivity analysis also involves the relationship between small changes in parameter values and the magnitudes of related changes in outputs. Such studies lead to a better understanding of how to structure the model with respect to the real system (Agin [1978], Emshoff and Sisson).

Given information about the underlying statistical distributions of model parameters, a Monte Carlo sensitivity analysis may lead to some useful statements about the distribution of the model outcomes. Sensitivity analysis is a way to compensate for uncertainty (Majone and Quade [1980]); it is a difficult and expensive enterprise and can involve advanced techniques such as response surface analysis (Harris [1980]).

For complex models, the most difficult of the operational validity concerns (and possibly in the validation area in general) is *implementation validity*. Here we are concerned with the extent to which the real world system being modeled responds in a manner indicated by the recommended model solution. For example, the model might specify that if a certain funding level is maintained, then a desired number of Medicare patients can be processed and sustained. Such a statement assumes that the human, organizational, and bureaucratic elements involved in the implementation will respond in the manner hypothesized by the model, for example, that case workers can and will, on the average, handle the case load level used by the model. Implementation validity of solutions to nonexisting systems must be considered in a different context than ongoing systems. Implementation validation for an operating system should be straightforward. We need to determine whether or not realistic plans exist that enable the solution to be implemented and imbedded within the real world problem environment. Implementation validity of a futuristic system is at best a judgmental affair.

Dynamic Validity

The final element of the validation framework is *dynamic validity*. Dynamic validity is concerned with determining how the model will be maintained during its life cycle so it will continue to be an acceptable representation of the real system. The two areas of interest are *updating* and *review*.

The model developers need to establish a procedure by which information is collected and analyzed to determine if and when model parameters or model structure need(s) to be changed, and a process by which such changes are affected. Also required is a review procedure, that is, a regular schedule for reviewing the success or failure of the model during its life cycle. These reviews must involve the decision makers, solution implementers, and model developers in sessions that cover divergences between the predicted solution and actual outcomes, proposed model changes, and a new evaluation of the model's validity.

Basic Validation Steps

There is no validation methodology appropriate for all models. As noted by Sisson [1975], we can never completely validate a decision-aiding model as there are never real data about the alternatives not implemented. Consequently, OR analysts must be extremely conscientious in devising, implementing, interpreting, and reporting validation tests for their models. We note that the basic validation steps of Emshoff and Sisson are still applicable:

(1) The analyst assures himself that the model performs the way he intends it to; using test data, and if available, real historical data.
(2) Reasonableness is checked by:
  (a) showing that key subsystem models predict their part of the world well (using historical data);
  (b) showing that ... the parameters have reasonable values;
  (c) having people who are knowledgeable about the situation (preferably including the decision maker) review the model in detail and agree to its structure and parameters.
(3) The decision maker has an opportunity to explore the use of the model to become familiar with its predictions and to examine the interactions it implies. At this point the analysts and decision maker may be able to agree as to what is a close enough fit between model output and actual data.
(4) The model is used for decision-aiding. Careful records are kept of its predictions and of actual results. (This may involve a time span of years, so that the evaluation procedure has to be set up carefully.)

More detailed approaches can be structured from the material described above (see Schellenberger, and Naylor and Finger [1967], and survey articles by Law [1979] and Sargent [1980]).

Validation Studies and Research

Papers describing validation attempts of specific applications can be found in the bibliographies by Deeter and Hoffman [1978] (on energy related models), and Balci and Sargent [1980] (on general simulation models). Stafford [1976] applies the technical validation steps of Schellenberger to a simulation model for multifacility outpatient clinics. Scheder [1976] using the definition "a simulation is validated when its users are satisfied with the output" describes a Monte Carlo based method for validating an air defense gun simulation model. Miller et al. [1976] extend the concept of sensitivity analysis to validating an ecological system simulation model. Schlesinger et al. [1974] apply their test procedures for reasonableness, validation, and verification to simulation models, and Hollis [1978] reports on approaches to validating a fiscal impact budgeting simulation. Research in the general validation area includes formalization of the process, statistical tests of validity, and approaches to sensitivity analysis. See, for example, Gruhl [1979], Mitchell and Wilson [1979], Fowler [1980], Balci and Sargent [1981], McKay [1979a, b, 1980], Schweppe and Gruhl [1980], Harris [1979], Greenberg and Murphy [1980], and Mayer [1980]. Validation tests for the class of system dynamics models are described in Forrester and Senge [1980].

The interest in energy models and their use in policy analysis has produced several relevant validation studies; see for example, Boshier et al. [1978], Singleton et al. [1979], Mitchell and Wilson, Marcuse et al. [1980], and Labys [1980]. Labys describes a number of statistical validation tests (parametric and nonparametric) and their applicability to the range of energy modeling techniques (econometric, input-output, mathematical programming, system dynamics). The Department of Energy and the National Bureau of Standards sponsored two symposia on energy model validation and assessment; the results are reported in Gass [1980, 1981a].

Validity questions have become more important now that OR models are used as aids in policy analysis. A key research task is the development of procedural and analytical methods for implementing the OR approach in the field of policy analysis. Whether this can be done—or is even appropriate—has been argued by several researchers (Strauch, Brewer [1975], Greenberger et al., House and McLeod, and Banfield [1980]). The OR community should rise to the challenge by pursuing a research agenda for improving the use of OR in policy analysis.

3. MODEL ASSESSMENT

Model assessment serves many purposes: education, model development and improvement, theoretical analysis, understanding the model and processes being studied, obtaining insights to aid policy analysis, data structure analysis, reproducibility of results, documentation, making the model more accessible, and determining the utility of the model for particular policy analyses. For this discussion, we shall view the main function of assessment to be a process by which interested parties (who were not involved in a model's origins, development, and implementation) can determine, with some level of confidence, whether or not the model's results can be used in decision-making (Gass [1977a]). There are three reasons for advocating assessment of policy models: (1) For many models, the ultimate decision maker is far removed from the modeling process (this is especially true in government) and needs a basis for deciding when to accept the model's results. (2) Users of a model developed for others must be able to obtain a clear statement of the applicability of the model to the new user problem area. (3) It is difficult to assess the impact of a model's assumptions, data availability, and other elements on the model structure and results without a formal, independent evaluation.

The OR analyst should be particularly concerned with the model evaluation process because the outcome of an evaluation is really a recommendation to the decision maker whether or not to use the output of the analyst's model. This is also the case for the computer programmer responsible for translation of the model into an operational computer program. OR and computer personnel must assume the responsibility for detailing their assumptions, simplifications, and methodologies, and for offering evidence to an evaluator that the rationale behind their approach will produce answers that can be used in the real-world environment as viewed by the ultimate decision maker.

Questionnaire for Evaluation

All procedures for assessing a model are basically information gathering activities, with the detail and level of information being a function of the purposes of the assessment and the skills of the assessors. If the purpose of the assessment is to determine the model's utility in a given policy study, the information must be obtained and interpreted. That is, the assessors must know under what circumstances the model will yield credible results and can be used with confidence. This is a difficult task and a judgmental one. Gass [1977b] proposes a paradigm to accomplish the evaluation in the form of a detailed "Questionnaire for the Utility Evaluation of a Complex Model." The questionnaire information is summarized by the assessors using a 13-item scale and corresponding rating scores. Based on user inputs, the items can be weighted and threshold values used to classify the model as (1) operationally acceptable and can be used with confidence, (2) acceptable with stated minor deficiences, or (3) not acceptable unless major deficiencies are rectified. The 13 information items are:

- Computer program documentation
- Model documentation
- Computer program consistency and accuracy
- Overall computer program verification
- Mathematical and logical description
- Technical validity
- Operational validity
- Dynamic validity
- Training
- Dissemination
- Usability
- Program efficiency
- Overall model validation.

MIT Energy Laboratory Assessment Approach

Personnel of the Massachusetts Institute of Technology (MIT) Energy Laboratory have evolved a multistep process for assessment, based on several energy models, with the steps being a function of the assessment purpose. See Figure 4 (Kresge [1979], Wood [1980]). The steps or elements of the MIT procedure are:

- **Review of the literature—an evaluation of a model's structure and characteristics which relies solely on the published materials. This first step of the assessment process brings together and summarizes the available information describing the model's objectives, struc-**

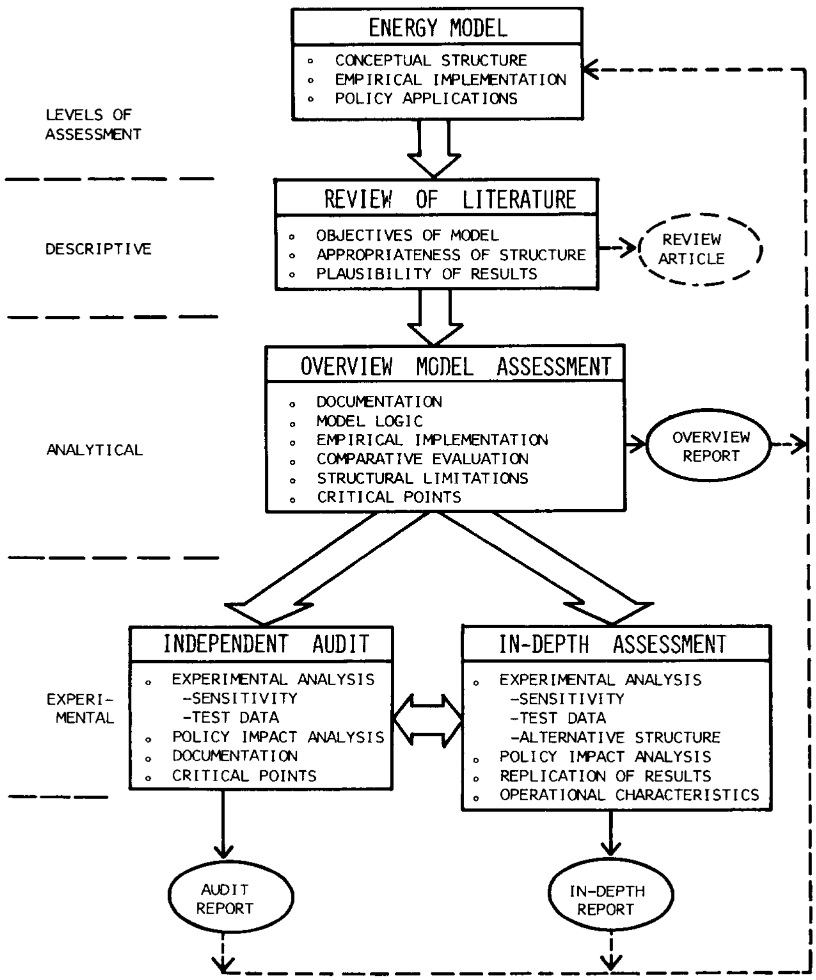

Figure 4. Approaches to energy model assessment. (From Kresge [1977].)

ture, and principal results. The review is primarily useful to a potential model user in providing a description of what the model is intended to do and of the methodology used to achieve the stated objectives.

- Overview model assessment—an analytical evaluation of the model's properties based on all (published and unpublished) technical documentation. It includes an evaluation of the empirical content of the model, limitations due to the model's structure, and identification of critical points and issues. It should include an analysis of the computer code. The overview report is an interim document used to plan the next steps of the assessment.
- Independent audit—an evaluation of the model's validity, applicability, and performance using data derived from experiments. The experiments are designed by the assessors but are implemented by the modelers with a member of the assessor team present. The audit report focuses on the model's behavior with regard to its major critical points, and provides information on the quality of the available documentation.
- In-depth assessment—an evaluation of the computer-based model by "hands-on" running by the assessor team. This allows for more complex tests, changes to the model by the assessor team, and identification of errors and discrepancies between implementation and documentation.

The above elements represent four distinct steps in model assessment, but they should be viewed as the stages in a comprehensive model assessment process. The extent to which an assessment covers all four elements is a function of the assessment purposes and, of course, personnel and financial resources. The result of an MIT assessment is to produce statements concerning the following items:

- Objectives of the model
- Appropriateness of the structure
- Plausibility of the results
- Documentation
- Model logic
- Empirical implementation
- Comparative evaluation
- Structural limitations
- Critical points
- Experimental analysis: sensitivity, test data, alternative structure
- Policy impact analysis
- Replication of results
- Operational characteristics.

The MIT procedure has been applied to the assessment of several energy models, see Goldman and Gruhl [1980], Kuh and Wood [1979], and Wood [1981]. Other assessment approaches are similar to the MIT procedure, with variations due to the requirements of the assessment and areas of expertise of the assessors; see Holloway [1980a,b], Weisbin et al. [1981], and Alsmiller [1980]. A general approach is given in U.S. GAO [1979].

Assessment Summary

Researchers in model assessment agree on one thing: the information required to perform an assessment should be produced by the model developers if they have adequate resources and time *and* if they adhere to basic professional OR guidelines. See Caywood et al. [1971], U.S. GAO [1976b], Brewer and Shubik, and Beraha [1981] for guidelines and model development procedures. Gass [1977a, 1981b] suggests that, if the modeling steps are accomplished in sufficient detail and scope, it should not be too difficult for independent assessors to estimate the model's validity and utility. The assessment could consist of replications and testing the veracity of the claims of the model developers. It is, of course, naive to think that a complex computer-based decision-aiding model is produced in working and resource conditions that yield a fully developed model ready for easy assessment. There are numerous problems. Specifically, the model is often a "moving target" that is modified continually (Wood [1980], Gass [1981b]). Analysts and programmers working in a policy environment are under pressure to produce answers and operate in a "crash mode" with little time to perform complete tests of the computer-based model (Greenberg [1979]). Model development guidelines or standards do not exist and, even if they did, it would be too expensive to adhere to them. An assessor cannot perform an assessment without involving the original model developers (Stauffer [1979]), and if this last claim is true, no model assessment can be truly independent.

The resolution of these points is really the resolution of what the assessment process should be. A model assessment procedure and its objectives should be tailored to the scope and purposes of the model and will vary with the model, model developers, assessors, users, available resources, etc. Model assessment is an expensive and involved undertaking; all models need not be assessed. Since policy analysis models will always be subject to some sort of a review, the initial development task should be of the highest professional caliber, one that allows the model to be readily assessed. This will save money in the long run, and be more useful to the policymaker in the short run. The preferred route to model assessment is to make a decision on the scope of the assessment and criteria for assessment when the model is started. In this way, the developers will know the ground rules and the material they must produce to satisfy the agreed upon criteria. It should be clear that more experimentation and research in assessment methodology is required to determine how to improve the utility of models for policy analyses.

4. RELATED ISSUES

Model Confidence and Credibility

Procedures for model validation and assessment tend to culminate in a judgment concerning the measure of confidence or credibility a user should give to the model's results (Emshoff and Sisson, Hoffman and Deeter [1979b], Gass and Joel [1980], and Sargent [1980]). At this time, no formal, precise definition and means of determining model confidence exists. We emphasize that model confidence is not an attribute of a model, but is an information-based opinion or judgment of a particular user for a given decision environment. The generation and presentation of the necessary information is the task of the model analysts and developers. The level of confidence may vary from user to user because of differences in application requirements, and subjective judgmental preferences. Confidence in a model evolves through a joint effort between the model developers and a designated user. For a decision model without a designated user (which implies specific use) there is no basis upon which a confidence statement can be made, that is, the *a priori* confidence level is zero. Certainly many analysts can demonstrate that their models give accurate predictions and have a high degree of validity, thus, the analysts have confidence in the model outputs. But for such models to be used in specific decision settings, they must be evaluated in terms of the decision maker's confidence-setting criteria. Such criteria are not necessarily quantitative. Whether or not a model satisfies a criterion depends on the analysts' (or assessors') abilities to produce specific information—model materials—required by the decision maker. In the ideal situation the decision maker and analysts agree to criteria and information needs prior to model development and testing. The final model materials should include the necessary information or explain why such information is unobtainable. A similar process should be part of any model assessment, since assessments make sense only if they are done for designated decision problems and hence, for an implied set of decision makers.

Model confidence is subjective, each decision maker internalizes the available information by means of an imprecise algorithm for evaluating model confidence. Such algorithms should be based on information produced by the initial model analysts or by subsequent model assessors. An approach to detailing information requirements and a demonstration of how model materials can be used by a decision maker to obtain a statement of model confidence is given in Gass and Joel.

Model Documentation

We do not know of any model assessment or modeling project review that indicated satisfaction with the available documentation. As documentation is the *sine qua non* of model validation and assessment, we cannot let the opportunity pass without noting that serious problems exist regarding the production and availability of model documentation. It is not our purpose to describe approaches for achieving adequate documentation, but to stress that model use, especially the use of policy analysis models, is a function of documentation. (See the surveys by Shubik and Brewer [1972], U.S. GAO [1974], and Fromm et al. [1975].) Readers interested in model and computer program documentation guidelines should refer to Brewer [1976], National Bureau of Standards (NBS) [1976, 1981], House and McLeod, Gass [1979a], Nance [1979], Newton and Weatherbee [1980], and Gass et al. [1981].

5. SUMMARY

The issue underlying the strong interest in policy-analysis model validation and assessment is our present inability to demonstrate how particular models can aid the policy process. This inability has been attributed to many things: lack of professionalism by the model developers, lack of support by the model sponsors, lack of faith by the model users, and possibly by demanding too much from the models. Decisions will be made with or without a supporting model. Every decision maker has a mental model against which the complex computer-based model is compared (Hogan [1978]). "... The fact remains that the computer model is normally the only one required to prove its credibility in any rigorous fashion before it is used to guide policymaking" (House and Ball). Hogan [1978], paraphrasing Geoffrion [1976] (who paraphrased Hamming [1962]), states that "The purpose of energy modeling is insight, not numbers," and "The challenge to the modeler and the modeling process is to use a complex model in a fashion which helps build this insight...." Maybe this is enough, or maybe this is all we can ask for, but the hope is to do more.

How to improve the utility of policy analysis models is the task for the OR community in the 1980s. The OR profession has a unique opportunity to extend its influence in this area due to the use of microcomputers and model-based Decision Support Systems as tools of the decision maker (see Keene and Morton [1978]). We can learn from computer scientists as they address the management of program development by structured programming procedures, the emergence of a software science to aid in program verification and reliability, documentation guidelines, and concerns with the quality of software. (See Hoare [1972], Fitzsimmons and Lowe [1978], NBS [1976], and Zelkowitz.) Ideas and approaches to OR modeling have not been lacking; witness the many references upon which this paper is based. Brewer [1977], Brewer and Shubik, Gass [1979b], Greenberg [1981], Greenberg and Murphy, U.S. GAO [1976b], and Ziegler

[1980] address the issues of professionalism, standards, and methods for model development and the management of the modeling process. House and McLeod offer "directions for progress" (p. 15). Renewed interest and directions in model validation and sensitivity analysis are indicated by McKay [1979a, b, 1980], Harris [1979, 1980], Labys, and Sargent. New approaches to model comparison and improvement procedures such as the Energy Modeling Forum are available (Hogan [1979], Sweeney and Weyant [1979], Sweeney [1981]). Many ideas are presented in Bayraktar et al. [1981], the validation and assessment workshop reports of Gass [1980, 1981b], and the conference proceedings of the sixth IIASA Symposium on Global Modeling by Meadows et al. [1982]. Greenberger et al. (Chapter 10 [1976]) describe approaches to improve the modeling and the political process. Richels [1981] presents a model analysis approach that bridges the often conflicting needs of analysts, assessors, and the policy makers. Greenberger [1981] offers steps that involve policy modelers and policymakers to humanize policy analysis and make policy models easier to use, understand, and accept for those without a technical orientation. And Gass and Parikh [1981] address the problem of how a multimodel analysis can be coordinated to produce a credible analysis.

The problems exist. The ideas are forthcoming. The need and the challenges are great.

ACKNOWLEDGMENTS

I wish to thank the referees and Bill Pierskalla for their helpful comments. Further acknowledgement is due to Bill for encouraging me to write this paper. Also, I want to express my appreciation for past support of my research in this area received from the Program Analysis Division, U.S. Government Accounting Office; The Energy Information Administration, Department of Energy; and the Operations Research Division, National Bureau of Standards. Finally, a review paper such as this one is based on the research of many people; I trust I have interpreted their work correctly. The reader is encouraged to consult the original sources, which include the following references that have served as general background for the material in this paper: Berman [1972], Crissey [1975, 1979], House [1974], Miller [1975], Nance [1981], Naylor [1973], Ören [1981], Pugh [1977], Sargent [1979], and Van Horn [1971].

REFERENCES

AGIN, N. I. 1978. The Conduct of Operations Research Studies, in Moder and Elmaghraby [1978].

ALSMILLER, R. G., JR. 1980. Interim Report on Model Evaluation Methodology and the Evaluation of LEAP, ORNL/TIM-7245, Energy Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

ALTMAN, S. M. 1980. Pitfalls of Data Analysis, in Majone and Quade [1980].

BALCI, O., AND R. G. SARGENT. 1980. Bibliography on Validation of Simulation Models, Spring 1980 Newsletter-TIMS College in Simulation and Gaming, pp. 11–15.

BALCI, O., AND R. G. SARGENT. 1981. A Methodology for Cost-Risk Analysis in Statistical Validation of Simulation Models. *Commun. ACM* **24,** No. 4, 190–197.

BANFIELD, E. C. (1980). Policy Science as Metaphysical Madness, in Goldwin [1980].

BAYRAKTAR, B. A., E. A. CHERNIAVSKY, M. A. LAUGHTON AND L. E. RUFF. 1981. *Energy Policy Planning.* Plenum Press, New York.

BERAHA, S. A. 1981. The Development of a Rational Framework for Policy Modeling, Ph.D. dissertation, Technion, Haiffa, Israel.

BERMAN, M. B. 1972. Notes on Validating/Verifying Computer Simulation Models, P-4841, The Rand Corporation, Santa Monica, Calif.

BOSHIER, J. F., F. C. SCHWEPPE AND J. GRUHL. 1978. Validity and Uncertainty of Predictive Energy Models. In *Proceedings of the Sixth Power Systems Computation Confidence*, Darmstadt, Germany.

BREWER, G. D. 1973. *Politicians, Bureacrats, and the Consultant.* Basic Books, New York.

BREWER, G. D. 1975. An Analyst's View of the Uses and Abuses of Modeling for Decisionmaking, P-5395, The Rand Corporation, Santa Monica, Calif.

BREWER, G. D. 1976. Documentation: An Overview and Design Strategy. *Simulation and Games* **7,** No. 3.

BREWER, G. D. 1977. Whatever Happened to Professionalism? paper presented at the ORSA/TIMS San Francisco meeting.

BREWER, G. D., AND M. SHUBIK. 1979. *The War Game.* Harvard University Press, Cambridge, Mass.

CAYWOOD, T. E., H. M. BERGER, J. H. ENGEL, J. F. MAGEE, H. J. MISER AND R. M. THRALL. 1971. Guidelines for the Practice of Operations Research. *Opns. Res.* **19,** 1123–1258.

CLARK, J., AND S. COLE. 1975. *Global Simulation Models: A Comparative Study.* John Wiley & Sons, New York.

CRISSEY, B. L. 1975. A Rational Framework for the Use of Computer Simulation Models in a Policy Context, Ph.D. dissertation, Johns Hopkins University.

CRISSEY, B. L. 1979. Models in the Policy Process: A Framework, in Gass [1979b], pp. 163–175.

DEETER, C. R., AND A. A. J. HOFFMAN. 1978. *An Annotated Bibliography of Mathematical Models Classified by Validation and Error Analysis Methods.* Texas Christian University, Fort Worth, Tex.

DEMILLO, R. A., R. J. LIPTON AND A. J. PERLIS. 1979. Social Processes and Proofs of Theorems and Programs. *Commun. ACM* **22,** No. 5, 271–280.

EMSHOFF, J. R., AND R. L. SISSON. 1970. *Design and Use of Computer Simulation Models.* Macmillan, New York.

FISHMAN, G. S., AND P. J. KIVIAT. 1967. Digital Computer Simulation: Statistical Considerations, RM-5387-PR, The Rand Corp., Santa Monica, Calif. (Also published as: The Statistics of Discrete Event Simulation, *Simulation*, April 1968).

FITZSIMMONS, A., AND T. LOWE. 1978. A Review and Evaluation of Software Science. *Comput. Surv.* **10,** No. 1.

FORRESTER, J. W., AND P. M. SENGE. 1980. Tests for Building Confidence in System Dynamics Models. *TIMS Studies Mgmt. Sci.* **14,** 209–228.

FOWLER, D. V. 1980. A Theoretical Structure for Analysis of Models, paper presented at the ORSA/TIMS Colorado Springs meeting, November 1980, SRI International, Menlo Park, Calif.

FROMM, G., W. L. HAMILTON AND D. E. HAMILTON. 1975. Federally Supported Mathematical Models—Survey and Analysis, U.S. GPO Stock No. 038-000-0021-0, Washington, D.C.

GASS, S. I. 1977a. Evaluation of Complex Models. *Comput. Opns. Res.* **4,** 27–35.

GASS, S. I. 1977b. A Procedure for the Evaluation of Complex Models. In *Proceedings First International Conference in Mathematical Models*, University of Missouri.

GASS, S. I. 1979a. Computer Model Documentation: A Review and Approach, National Bureau of Standards (NBS) Special Publication 500-39, Washington, D.C.

GASS, S. I. (ed.) 1979b. Utility and Use of Large-Scale Mathematical Models, Proceedings of a Workshop, NBS Special Publication 534, Washington, D.C.

GASS, S. I. (ed.) 1980. Validation and Assessment Issues of Energy Models, Proceedings of a Workshop, NBS Special Publication 564, Washington, D.C.

GASS, S. I. (ed.) 1981a. Validation and Assessment of Energy Models, Proceedings of a Symposium, NBS Special Publication 616, Washington, D.C.

GASS, S. I. 1981b. Validation and Assessment Issues of Energy Models, pp. 421–441 in Bayraktar et al. [1981].

GASS, S. I., AND L. S. JOEL. 1981. Concepts of Model Confidence. *Comput. Opns. Res.* **8,** No. 4, 341–346.

GASS, S. I., AND S. C. PARIKH. 1981. Credible Baseline Analysis for Multi-Model Public Policy Studies. In *Proceedings IFIP Working Conference on Global Models*, Dubrovnik, September 1-5, 1980.

GASS, S. I., AND R. L. SISSON (eds.) 1975. *A Guide to Models in Governmental Planning and Operations.* Sauger Books, Potomac, Md.

GASS, S. I., AND B. W. THOMPSON. 1980. Guidelines for Model Evaluation: An Abridged Version of the U.S. General Accounting Office Exposure Draft. *Opns. Res.* **28,** 431–439.

GASS, S. I., K. L. HOFFMAN, R. H. F. JACKSON, L. S. JOEL AND P. B. SAUNDERS. 1981. Documentation for A Model: A Hierarchical Approach. *Commun. ACM* **24,** No. 11, 728–733.

GEOFFRION, A. M. 1976. The Purpose of Mathematical Programming Is Insight, Not Numbers. *Interfaces* **7,** No. 1, 81–92.

GOLDMAN, N. L., AND J. GRUHL. 1980. Assessing the ICF Coal and Electric Utilities Model, in Gass [1980].

GOLDWIN, R. A. (ed.) 1980. *Bureaucrats, Policy Analysts, Statesmen: Who Leads?* American Enterprise Institute, Washington, D.C.

GOODE, H. H. 1957. The Application of a High Speed Computer to the Definition and Solution of the Vehicular Traffic Problem. *Opns. Res.* **5,** 775–793.

GREENBERG, H. J. 1979. The FEA Protect Independence Experience, pp. 111–122, in Gass [1979b].

GREENBERG, H. J. 1981. A Tutorial on Computer-Assisted Analysis. In *Advanced Techniques for Operations Research in Practice*, H. Greenberg, F. Murphy and S. Shaw (eds.). Elsevier-North Holland, New York.

GREENBERG, H. J., AND F. MURPHY. 1980. Validity is a Composite Measure of Goodness, in Gass [1980].

GREENBERGER, M. 1981. Humanizing Policy Analysis: Confronting the Paradox in Energy Policy Modeling, in Gass [1981a].

GREENBERGER, M., M. A. CRENSON AND B. L. CRISSEY. 1976. *Models in the Policy Process.* Russell Sage Foundation, New York.

GREENBERGER, M., AND R. RICHELS. 1979. Assessing Energy Policy Models: Current State and Future Directions. *Ann. Rev. Energy*, **4**.

GRIFFIN, D. S. 1970. The Verification and Acceptance of Computer Programs for Design Analysis. In *On General Purpose Finite Element Computer Programs*, pp. 143–150, P. V. Marcel (ed.). ASME, New York.

GRIFFIN, D. S. 1971. The Qualifications of Solution Methods. In *Proceedings of the Symposium on Engineering Computer Software*, ASME, New York.

GRUHL, J. 1979. Model Validation. In *Proceedings of the International Conference on Cybernetics and Society*, IEEE Catalog No. 79CH1424-1SMC, Denver, Col.

HAMMING, R. W. 1962. *Numerical Methods for Scientists and Engineers*. McGraw-Hill, New York.

HARRIS, C. F. 1979. Model Validation, Simulation, and Sensitivity Analysis, CMH2-NBS-DOE, National Bureau of Standards, Washington, D.C.

HARRIS, C. F. 1980. A Statistical Method for the Assessment of Model Sensitivity to Input Variables, Project Report CMH3-NBS-DOE, National Bureau of Standards, Washington, D.C.

HERMANN, C. 1967. Validation Problems in Games and Simulations. *Behavioral Sci.* **12,** 216–230.

HILLIER, F., AND G. LIEBERMAN. 1980. *Operations Research*, Ed. 3. Holden-Day, San Francisco.

HIRSHFELD, D. 1980. Investigation of Underlying Data: Midterm Oil and Gas Supply Modeling System, Project Report KFR 245-79, National Bureau of Standards, Washington, D.C.

HOARE, C. A. R. 1972. Guest Editorial: The Quality of Software. *Software-Practice and Experience* **2,** 103–105.

HOFFMAN, A. A. J., AND C. R. DEETER. 1979a. *Proceedings of the Workshop in Validation of Computer-Based Mathematical Models in Energy Related Research and Development*, Texas Christian University, Fort Worth, Tex.

HOFFMAN, A. A. J., AND C. R. DEETER. 1979b. Final Draft of the Workshop on Validation of Computer-Based Mathematical Models in Energy Related Research and Development, Texas Christian University, Fort Worth, Tex.

HOGAN, W. W. 1978. Energy Modeling: Building Understanding for Better Use, paper presented at the 2nd Lawrence Symposium on The Systems and Decision Sciences, Berkeley, Calif., October 3, 1978.

HOGAN, W. W. 1979. The Energy Modeling Forum, pp. 137–161, in Gass [1979b].

HOLLIS, M. S. 1978. Methodological Issues in the Validation of Fiscal Impact Budgeting Simulation Models, paper presented at the ORSA/TIMS Los Angeles Meeting, November 1978, Graduate School of Administration, University of California, Irvine, Calif.

HOLLOWAY, M. L. 1980a. The Texas National Energy Modeling Project: An Evaluation of EIA's Midrange Energy Forecasting System, pp. 211–236, in Gass [1980].

HOLLOWAY, M. L. (ed.) 1980b. *Texas National Energy Modeling Project: An Experiment in Large-Scale Model Transfer and Evaluation*, Parts I, II, and III. Academic Press, New York.

HOOS, I. R. 1972. *Systems Analysis in Public Policy*. University of California Press, Berkeley, Calif.

HOUSE, P. W. 1974. Diogenes Revisited—The Search for a Valid Model. *Simulation*, pp. 117–125 (October).

HOUSE, P. W., AND R. H. BALL. 1980. Validation: A modern Day Snipe Hunt? Conceptual Difficulties of Validating Models, in Gass [1980].

HOUSE, P. W. AND J. MCLEOD. 1977. *Large-Scale Models for Policy Evaluation*. John Wiley & Sons, New York.

KEENE, P. G., AND M. S. MORTON. 1978. *Decision Support Systems: An Organizational Perspective*. Addison-Wesley, Reading, Mass.

KRESGE, D. 1979. The EPRI/NBER Energy Model Assessment Project, pp. 123–135, in Gass [1979b].

KUH, E., AND D. O. WOOD. 1979. Independent Assessment of Energy Policy Models, Final Draft, Research Project 1015-1, MIT Energy Laboratory, Cambridge, Mass.

LABYS, W. C. 1980. Measuring the Validity and Performance of Energy Models, College of Mineral and Energy Resources, West Virginia University, Morgantown, W.V.

LARSON, R. C., AND A. R. ODONI. 1981. *Urban Operations Research*. Prentice-Hall, Englewood Cliffs, N.J.

LAW, A. M. 1979. Validation of Simulation Models, I: An Overview and Survey of Real-World Practice, Working Paper 78-14, Department of Industrial Engineering, University of Wisconsin-Madison.

LAW, A. M., AND K. KELTON. 1981. *Stimulation Modeling and Analysis*. McGraw-Hill, New York.

LEDBETTER, W. N., AND J. F. COX. 1977. Are OR Techniques Being Used? *Industrial Engineering*, February.

LEE, D. B. 1973. Requiem for Large-Scale Models. *J. Am. Inst. Planners* **39**, 163–178.

MAJONE, G., AND E. S. QUADE, (eds.) 1980. *Pitfalls of Analysis*. John Wiley & Sons, New York.

MARCUSE, W., F. T. SPARROW, AND D. A. PILATI. 1980. Validation Issues: A View from the Trenches, in Gass [1980].

MAYER, L. S. 1980. On a Perspective for Energy Model Validation, in Gass [1980].

MCKAY, M. D. 1979a. Sensitivity Analysis, in Hoffman and Deeter [1979a].

MCKAY, M. D. 1979b. Sensitivity Analysis and a National Energy Model Example, LA-UR-79-2895, Los Alamos Scientific Laboratory, Los Alamos, N.M.

MCKAY, M. D. 1980. Thoughts on Sensitivity Analysis and Variation, paper presented at the ORSA/TIMS Colorado Springs meeting, November 1980, LA-UR-80-3223, Los Alamos Scientific Laboratory, Los Alamos, N.M.

MEADOWS, D., J. RICHARDSON AND G. BRUCKMANN. 1982. *Groping in the Dark.* John Wiley & Sons, London.

MILLER, D. R. 1975. Recent Developments in Approaches to Validation of Simulation Models, paper presented at the ORSA/TIMS Las Vegas meeting, November 1975, National Research Council of Canada, Ottawa, Canada.

MILLER, D. R., G. R. BUTLER AND L. BRAMELL. 1976. Validation of Ecological System Models. *J. Environ. Mgmt.* **4,** 383–401.

MITCHELL, T. J., AND D. G. WILSON. 1979. Energy Model Validation: Initial Perceptions of the Process, ORNL/CSD-50, Oak Ridge National Laboratory, Oak Ridge, Tenn.

MODER, J. J., AND S. E. ELMAGHRABY. 1978. *Handbook of Operations Research.* Van Nostrand Reinhold, New York.

NANCE, R. E. 1979. Model Representation in Discrete Event Simulation: Prospects for Developing Documentation Standards. In *Current Issues in Computer Simulation*, pp. 83–87, N. R. Adam and A. Dogramaci (eds.). Academic Press, New York.

NANCE, R. E. 1981. Model Representation in Discrete Event Simulation: The Conical Methodology, Tech. Report CS81003-R, Dept. of Computer Science, Virginia Polytechnic Institute and State University.

NAYLOR, T. H. 1973. Simulation and Validation. In *OR '72*, pp. 205–222, M. Ross (ed.). North-Holland, Amsterdam.

NAYLOR, T. H., AND J. M. FINGER. 1967. Verification of Computer Simulation Models. *Mgmt. Sci.* **14,** B92–B101.

NATIONAL BUREAU OF STANDARDS (NBS) 1976. Guidelines for Documentation of Computer Programs and Automated Data Systems, Federal Information Processing Standard (FIPS) Publication 38, National Bureau of Standards, Washington, D.C.

NBS 1981. Computer Model Documentation Guide, NBS SP 500-73, S/N 603-003-02282-9, U.S. Government Printing Office, Washington, D.C.

NEWTON, O. L., AND J. E. WEATHERBEE. 1980. Guidelines for Documenting Computer Simulation Models. In *Proceedings of the 1980 Winter Simulation Conference*, Orlando, Fla.

ORDEN, A. 1979. The Effectiveness of Linear Programming Models in Operations Planning, paper presented at the 10th International Symposium on Mathematical Programming, Montreal, August 27–31.

ÖREN, T. I. 1981. Concepts and Criteria to Assess Acceptability of Simulation Studies: A Frame of Reference. *Commun. ACM* **24,** No. 4, 180–189.

PROFESSIONAL AUDIT REVIEW TEAM (PART) 1977. Activities of the Office of Energy Information and Analysis, U.S. General Accounting Office, Washington, D.C., November 5.

PART 1979. Activities of the Energy Information Administration, U.S. General Accounting Office, Washington, D.C., May 7.

PECK, J. R. 1974. The Use of Urban Models in Urban Policy Making, The Fels Center of Government, University of Pennsylvania.

PUGH, R. E. 1977. *Evaluation of Policy Simulation Models.* Information Re-

sources Press, Washington, D.C.
QUADE, E. S. (ed.) 1966. *Analysis for Military Decisions.* Rand McNally, Chicago.
QUADE, E. S. 1980. Pitfalls in Formulation and Modeling, in Majone and Quade [1980].
QUADE, E. S. 1982. *Analysis for Public Decisions*, Ed. 2. Elsevier, New York.
RICHELS, R. 1981. Building Good Models is Not Enough, in Gass [1981].
SARGENT, R. G. 1979. Validation of Simulation Models. In *Proceedings of the 1979 Winter Simulation Conference*, San Diego, Calif.
SARGENT, R. G. 1980. Verification and Validation of Simulation Models, Working Paper No. 80-013, College of Engineering, Syracuse University, Syracuse, N.Y.
SCHEDER, R. A. 1976. Model Validation—A Case Study, Aberdeen Proving Ground, Aberdeen, Md.
SCHELLENBERGER, R. E. 1974. Criteria for Assessing Model Validity for Managerial Purposes. *Decision Sci.* **5**, No. 4, 644–653.
SCHLESINGER, S., ET AL. 1974. Developing Standard Procedures for Simulation Validation and Verification. In *1974 Summer Computer Conference Proceedings*, pp. 929–933.
SCHWEPPE, F. C., AND J. GRUHL 1980. Systematic Sensitivity Analysis Using Describing Functions, in Gass [1980].
SHANNON, R. E. 1975. *Systems Simulation: The Art and Science.* Prentice-Hall, Englewood Cliffs, N.J.
SHUBIK, M., AND G. BREWER. 1972. Models, Simulations, and Games, The Rand Corp., Report R-1060-A12PA/RC, Santa Monica, Calif.
SINGLETON, F. D., JR., S. MUTHUKRISHNAN, AND R. G. THOMPSON. 1979. Verification/Validation of an Economic Process Model for Electric Power Generation in the United States. *Comput. Opns. Res.* **8**, No. 4, 311–339.
SISSON, R. L. 1975. Introduction to Decision Models, in Gass and Sisson [1975].
STAFFORD, E. F., JR. 1976. Technical Validation of a Simulation Model, paper presented at the ORSA/TIMS Miami Meeting, November 4, 1976, University of Oklahoma, Stillwater, Okla.
STAUFFER, C. H., JR. 1979. Developing, Improving, and Assessing the ICF Coal and Electric Utilities Model, pp. 141–151 in Gass [1980].
STRAUCH, R. E. 1974. A Critical Essay of Quantitative Methodology as a Policy Analysis Tool, The Rand Corp., P-5282, Santa Monica, Calif.
SWEENEY, J. L. 1981. Model Comparison for Energy Policy and Planning, pp. 259–285 in Bayraktar et al. [1981].
SWEENEY, J. L., AND J. P. WEYANT. 1979. The Energy Modeling Forum: Past, Present, and Future, EMF PP 6.1, The Energy Modeling Forum, Stanford University, Stanford, Calif.
U.S. GENERAL ACCOUNTING OFFICE (GAO) 1971. Computer Simulations, War Gaming, and Contract Studies, B-163704, Washington, D.C.
U.S. GAO 1974. Improvement Needed in Documenting Computer Systems, B-115369, Washington, D.C.
U.S. GAO 1976a. Review of the 1974 Project Independence Evaluation System, OPA-76-20, Washington, D.C.
U.S. GAO 1976b. "Ways to Improve Management of Federally Funded Computerized Models, LCD-75-111, U.S. General Accounting Office, Washington, D.C.
U.S. GAO 1977. An Evaluation of the Use of the Transfer Income Model—TRIM—to Analyze Welfare Programs, PAD-78-14, Washington, D.C.

U.S. GAO 1979. Guidelines for Model Evaluation, PAD-79-17, Washington, D.C.

U.S. LIBRARY OF CONGRESS. 1975. Computer Simulation Methods to Aid National Growth Policy. Prepared by Futures Research Group for the Subcommittee on Fisheries and Wildlife Conservation and the Environment. U.S. Government Printing Office, Washington, D.C.

VAN HORN, R. L. 1971. Validation of Simulation Results. *Mgmt. Sci.* **17,** 247–258.

WALKER, W. E., J. M. CHAIKEN AND E. J. IGNALL (eds.) 1979. *Fire Department Deployment Analysis: A Public Policy Analysis Case Study.* Elsevier/North-Holland, New York.

WEISBIN, C. R., R. W. PEELE AND A. S. LOEBL. 1981. An Approach to Evaluating Energy-Economy Models, ORNL-5742, Engineering Physics Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

WOOD, D. O. 1980. Model Assessment and The Policy Research Process: Current Practice and Future Promise, pp. 23–62, in Gass [1980].

WOOD, D. O. 1981. Energy Model Evaluation and Analysis: Current Practice, in Gass [1981].

ZELKOWITZ, M. V. 1978. Perspectives in Software Engineering. *Comp. Surv.* **10,** No. 2.

ZIEGLER, B. P. 1976. *Theory of Modelling and Simulation.* John Wiley & Sons, New York.

ZIEGLER, B. P. 1980. Concepts and Software for Advanced Simulation Methodologies. In *Proceedings of the 1980 Winter Simulation Conference*, Vol. 2, pp. 25–44, T. I. Ören, C. M. Shub and P. F. Roth (eds.).