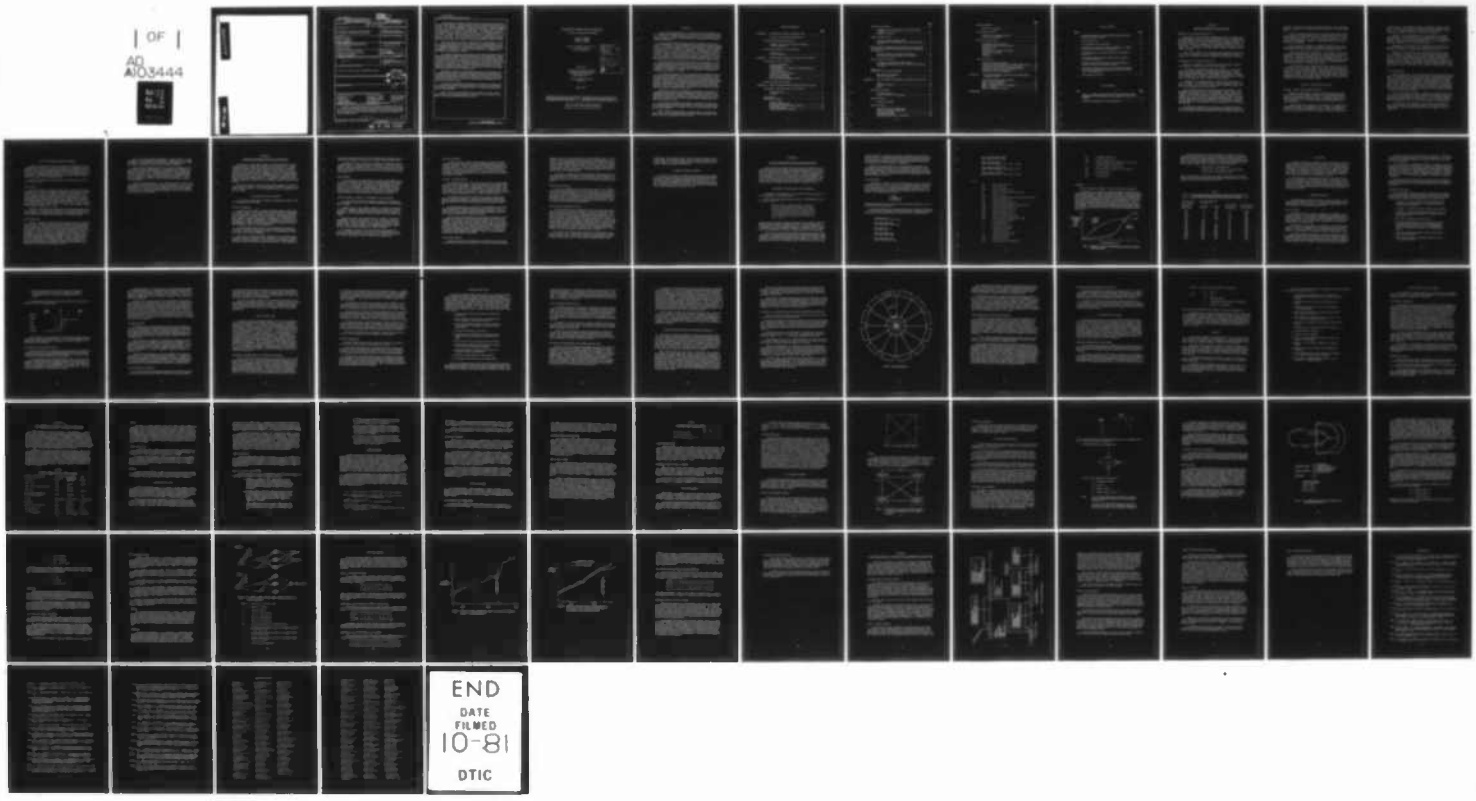
AD-A103 444 APPLIED PSYCHOLOGICAL SERVICES INC WAYNE PA SCIENCE --ETC F/G 9/2
EVALUATIONS OF OPERATIONAL DECISION AIDS. 3. EVALUATIVE APPROAC--ETC(U)
AUG 80 A I SIEGEL, E G MADDEN, M G PFEIFFER N00014-77-C-0048

UNCLASSIFIED NL



(1) Final rept. Jul 77-Aug 81

LEVEL II

A089127

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A103 444 / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Evaluations of Operational Decision Aids: 3. Evaluative Approaches and Methods	5. TYPE OF REPORT & PERIOD COVERED Final July 1977-August 1981 / 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Arthur I. Siegel, Edward G. Madden, Mark G. Pfeiffer	8. CONTRACT OR GRANT NUMBER(s) N0014-77-C-0448
9. PERFORMING ORGANIZATION NAME AND ADDRESS Applied Psychological Services, Inc. Science Center Wayne, PA 19087	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-018/5-16-80 455
11. CONTROLLING OFFICE NAME AND ADDRESS Engineering Psychology Programs Office of Naval Research Arlington, VA 22217	12. REPORT DATE (11) August 1980 / 13. NUMBER OF PAGES (12) 72
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (15) N0014-77-C-0448	15. SECURITY CLASS. (of this report) UNCLASSIFIED / 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE AUG 28 1981 S H

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

85

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Evaluation Research	Experimental Methods	Quasiexperiments
Methodology	Information Systems	Analytic Methods
Decision Aiding	Information Analysis	System Analysis
Decision Making	Aid Evaluation	User Test
Information Processing	Management Decision Making	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Concepts and considerations in the evaluation of operational decision aids are reviewed, and specific approaches to the evaluation of such aids are presented. The result is a unified, critical compendium of use to the developers of operational decision aids, outside evaluation agencies, and decision aid development progress administrators.

4/02/74 JOB

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

80 8 28 025

The purposes of operational decision aids are presented and their characteristics are described. Then, operational decision aid evaluation criteria are discussed. This discussion includes proximal vs. distal criteria, criterion acceptability and how to assure acceptability, criterion generalizability, criterion reliability, criterion objectivity, and criterion quantifiability. The problem of the "practical significance" of a criterion in terms of provision of operationally oriented data is treated. Special criterion problems such as internal vs. external criteria, locus of the evaluation, and post facto derived criteria are discussed vis-a-vis decision aid evaluation. The use of multiple criteria is suggested because multiple measures which support one another are less vulnerable than a unitary measure.

The major content is concerned with methods for aid evaluation: (1) analytic (paper-and-pencil and interview) methods, (2) experimental methods, and (3) quasiexperimental methods. Within the discussion of each method, a number of techniques is considered. Each technique is elaborated to demonstrate how it is applicable to decision aid evaluation, and the strengths and weaknesses are discussed, as appropriate, in context.

The analytic techniques are largely paper-and-pencil oriented and may be employed to describe selected traits or characteristics of decision aids by following reductionistic methods. They assume that an aid can be evaluated in terms of its elements and that the benefits of an aid can be understood by studying the elements. Cost Analysis, the Polydiagnostic Method, the Analytic Profile System, the Display Evaluation Index, the Perceptual Organization and Reduction Questionnaire, multiattribute utility analysis, and interview methods are separately discussed, reviewed, and commented on. Measurement considerations and application generality are treated as related issues.

Experimental methods are dichotomously discussed as "experimental" and "quasiexperimental" with the principal difference being whether the evaluator manipulates an independent variable or naturalistic variation is allowed. Formal experimental design is only briefly presented because this topic is covered in a wide variety of current texts. However, special design considerations for decision aid evaluation are given.

An overview of quasiexperimental designs is presented, and methods for treating quasiexperimental data are elaborated. These include: partial correlation, multiple regression, cross-lagged correlation, structural equation models, and time series analysis.

Finally, decision aid development is presented within a system development context, and the evaluative methods which seem most appropriate at each stage of the system developmental cycle are given.

EVALUATIONS OF OPERATIONAL DECISION AIDS

3. Evaluative Approaches and Methods

Arthur I. Siegel
Edward G. Madden
Mark G. Pfeiffer

Applied Psychological Services, Inc.
Science Center
Wayne, Pennsylvania

prepared for

Engineering Psychology Programs
Psychological Sciences Division
Office of Naval Research
Washington, D.C.

Contract N00014-77-C-0448
NR198-018

August 1981

This work formed a part of the operational decision aid program of the Office of Naval Research.

ABSTRACT

Concepts and considerations in the evaluation of operational decision aids are reviewed, and specific approaches to the evaluation of such aids are presented. The result is a unified, critical compendium of use to the developers of operational decision aids, outside evaluation agencies, and decision aid development progress administrators.

The purposes of operational decision aids are presented and their characteristics are described. Then, operational decision aid evaluation criteria are discussed. This discussion includes proximal vs. distal criteria, criterion acceptability and how to assure acceptability, criterion generalizability, criterion reliability, criterion objectivity, and criterion quantifiability. The problem of the "practical significance" of a criterion in terms of provision of operationally oriented data is treated. Special criterion problems such as internal vs. external criteria, locus of the evaluation, and post facto derived criteria are discussed vis-a-vis decision aid evaluation. The use of multiple criteria is suggested because multiple measures which support one another are less vulnerable than a unitary measure.

The major content is concerned with methods for aid evaluation: (1) analytic (paper-and-pencil and interview) methods, (2) experimental methods, and (3) quasiexperimental methods. Within the discussion of each method, a number of techniques is considered. Each technique is elaborated to demonstrate how it is applicable to decision aid evaluation, and the strengths and weaknesses are discussed, as appropriate, in context.

The analytic techniques are largely paper-and-pencil oriented and may be employed to describe selected traits or characteristics of decision aids by following reductionistic methods. They assume that an aid can be evaluated in terms of its elements and that the benefits of an aid can be understood by studying the elements. Cost Analysis, the Polydiagnostic Method, the Analytic Profile System, the Display Evaluation Index, the Perceptual Organization and Reduction Questionnaire, multiattribute utility analysis, and interview methods are separately discussed, reviewed, and commented on. Measurement considerations and application generality are treated as related issues.

Experimental methods are dichotomously discussed as "experimental" and "quasiexperimental" with the principal difference being whether the evaluator manipulates an independent variable or naturalistic variation is allowed. Formal experimental design is only briefly presented because this topic is covered in a wide variety of current texts. However, special design considerations for decision aid evaluation are given.

An overview of quasiexperimental designs is presented, and methods for treating quasiexperimental data are elaborated. These include: partial correlation, multiple regression, cross-lagged correlation, structural equation models, and time series analysis.

Finally, decision aid development is presented within a system development context, and the evaluative methods which seem most appropriate at each stage of the system developmental cycle are given.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

CONCEPTUALIZATION OF DECISION AIDS

Definition of an Operational Decision Aid

This report is about operational decision aids and how to evaluate them. Operational decision aids are found in many different forms and possess varying characteristics. However, some implications concerning the meaning of the term can be abstracted. To "aid" means to sustain, to help, or to assist. In the present context, the assistance is to the decision making activity of a decision maker. A decision aid helps the decision maker to bring to an end vacillation, doubt, and dispute by clarifying choices relative to judgments which lead to an action.

In the present context, we are concerned with *operational* decision aids in the Navy. Accordingly, we are concerned with decision aids which are relevant to Navy actions or missions, including their planning and execution.

Characteristics of Operational Decision Aids

Within such a definition, operational decision aids possess a number of characteristics--at least as they are currently composed.

The aids are usually computer based and interactive. The computer basing of operational decision aids implies other characteristics. One of these is data base dependency. This dependency may represent a limiting characteristic because the quality of the decision aid's output is, to some extent, conditioned by the nature of the information in the data base, its scope, accuracy, dependability, and how easily it can be updated.

A decision aid's interactive nature is a token of the fact that it is an "aid" or "helper." It is intended to allow a user (the decision maker) to interact dynamically with the aid in order to explore alternative courses of action or the consequences of various decisions.

Aids generally possess a rather complex internal architecture. This complexity is an artifact of and, in many instances, a mirror of the highly diversified considerations inherent in the decision(s) the aid supports. For example, a Navy decision about how to deploy its own resources must also consider items such as the enemy's strength, disposition, intent, and logistic situation along with weather and possible political repercussions. And, each of these factors does not operate in isolation. Each interacts with others. The interactions are often impossible to perceive beforehand or to understand after the fact.

The internal architecture of an aid acts as a high fidelity sounding board which applies the same rules and algorithms to different input information and allows the potential consequences of different courses of action to be

examined. Obviously, the amount of assistance provided by this contingency examination depends on the veridicality and sophistication of the aid's internal logic.

Another characteristic that seems to be shared by most aids is flexibility of input to the user's judgment. Most advanced decision aids allow personal probabilities, weights, or values to be entered by the user. The user is given some discretion in the entry value(s) of input parameters so as to allow the user's judgment, experience, or subjective values to be incorporated into the calculation.

At the simplest level, weights are employed to modulate scale values and influence responses, judgments, or utilities during the aid's calculation. Linear models are normally used to implement this modulatory action.

A related characteristic of decision aids is their reliance on some type of simulation model within their internal logic. Unfortunately, many of the models fail to consider all important terms or factors within the total area of interest. Undoubtedly, this facilitates the development of the aids. However, the fact that important factors have not been modeled may exact a cost. Important first order and interaction effects may not be considered in the model's algorithms with a resultant barren output.

Another characteristic of decision aids is their feedback function. The purpose of an aid is to provide feedback to the user, who must make some sort of judgment. The feedback must at least sensitize the user to the consequences of various courses of action. The importance of the feedback mechanism can be gauged from the Madden and Siegel (1980) study of an aid developed for the Navy. They reported that the perceived utility of the aid, obtained from a multiattribute utility analysis, appeared to vary directly with the adequacy of the aid's feedback functions.

The Role of Operational Decision Aids

Strategic, Tactical, and Contingency Planning

Depending on their design, purpose, and interface structure, aids can support the decision maker in his planning and analysis for strategic, tactical, and contingency planning purposes. Although all aids are not equally applicable to all areas, most seem to be specifically designed to assist, regardless of area, in assessing tradeoffs between factors of importance to the users.

Strategic planning and analysis can be supported in a number of ways. Planning begins with a clear statement of goals. Often, the goals can be achieved in a number of different ways, e.g., a mission to neutralize an enemy task force might be achieved by blockade, air strike, surface attack,

or missile attack. Each possibility requires very different massing and commitment of forces. Aids such as the decision structuring process aid (Markhofer, et al., 1979) are designed to facilitate this type of planning. They support the development and analysis of alternative strategies.

Similarly, operational decision aids can be instrumental in assisting tactical planning. In such planning, there are intricate and interrelated judgments which can benefit from aiding applications, e.g., the composition, ordinance, and armament mix of an attacking force. Here, the tradeoff and analysis capabilities of an aid are particularly useful.

Contingency planning represents a combination of both strategic and tactical considerations and factors. The feedback function and interactive aspects of aids support their application to contingency planning.

Related to strategic, tactical, and contingency planning is the problem of the timing of an action. Once a course of action has been decided on, the optimal time to implement the action often becomes a crucial consideration. To date, there have been few aids that attempt to address the timing question. Analytics' strike timing decision aid (Glenn & Zachary, 1978) represents one aid devoted to this purpose. Aiding judgments about timing problems was found to possess some degree of utility (Siegel & Madden, 1980), and it may prove that, in the future, the concept will be applied more frequently.

Training Applications

While the ultimate use of an *operational* decision aid is during actual operations, the advantages that such aids can afford to training should not be overlooked. However, as pointed out by Sinaiko (1976), the training advantages may not be a sufficient end in themselves. Sinaiko suggested that the training application of aids must be tempered by specifying who is to be trained and for what. Certainly, training can occur at many levels and for many purposes. At the highest level, the idea of using a decision aid to support fleet or task force level operations training seems to represent a suitable application. Aids can be employed, for example, in this context to demonstrate the effects of variables on outcomes and payoffs.

The training capability inherent in decision aids is supported by the results of a set of interviews with officer personnel who employed a decision aid for research purposes (Madden & Siegel, 1980). The interviewees indicated several training applications of the aid involved: (1) as a training device for task force commanders, and (2) as a tool for training in plan development.

As the result of such applications, the user would gain insight into areas to be considered in formulating judgments, errors within his own decision process, tradeoff potentials relative to various situations, and risk minimization.

Role of Evaluation During Aid Development

The body of the present report is concerned with methods for evaluating operational decision aids. The need for objective, quantitative test and evaluation of such aids was emphasized at three levels by Sinaiko (1976): (1) initial in-house test by the persons responsible for the development of an aid, (2) test by an independent outside agency, and (3) test by an appropriate user agency. This hierarchical evaluation process allows critical and fair assessment throughout the developmental cycle.

The various levels may be conceptualized as internal (level 1) and external (levels 2 and 3).

Internal Level

Internal evaluation of operational decision aids has been used with considerable success relative to a number of objectives. These include: determining features to be added to or deleted from aid prototypes (Barclay et al., 1979), selection of interactive graphic control devices (Irving et al., 1976), model development (Walsh & Schechterman, 1976), testing alternative forms of optimzing functions (Schechterman & Walsh, 1980), estimating the correspondence between model output and "expert" judgment (Glenn, 1978), use of option selection matrices to support decision aids (Kalenty et al., 1977), demonstrating adaptive decision aiding (Leal et al., 1978), and extending a taxonomy of Naval decisions (Miller et al., 1980).

Generally, these studies have allowed for insight into possible development, refinement, and adjustment of aids during the development stage. They can be conceived as enabling the development of substantially suitable operational decision aids and making the aids ready for external evaluation.

External Evaluation

External evaluations are performed by some agency other than the developer of the aid. Basically, external evaluations attempt to answer questions about the merit of an operational decision aid at a variety of analytic levels. At the highest level, the questions concern the general usefulness of an aid. To what extent does it allow a user to make "better" decisions? Under what conditions? The issues raised by such questions reflect the heart and central purpose of an aid. Because of the importance of the answer to such questions, rigorous methodologies are employed during the evaluation including direct manipulation of independent variables, control of error sources, and precise measurement of response variables. While less rigorous techniques are often included as a supplement within external evaluations, the more rigorous approach is the fundamental ingredient because the external evaluation often represents the final test of an aid.

We do not conceive external evaluation as a unitary process. It may take place in steps proceeding from laboratory, through test in a high fidelity simulator, and culminate in a fleet technical and operational test. Less rigor may be possible in the latter types of external evaluation.

From a somewhat different point of view, external evaluation may also be concerned with the instrumentality of the components of a decision aid. What is it about the aid that facilitates its usefulness? What elements are used by the decision maker? How consistently? Can a causal relationship be observed between a user's decisions and selected characteristics of the aid? To answer such questions, observational data are acquired. These data are employed in a variety of ways to develop models and to evaluate the user's policy.

From a still different point of view, external analysis can answer questions related to the perceived utility and goal attainment of an aid. Do users perceive the aid to have utility? How closely are the goals of the decision aid achieved? Is all required information presented? How well? How easy is the aid to use? To what extent would users employ it? How much confidence do users have in the aid's output?

CHAPTER II

OPERATIONAL DECISION AID EVALUATION CRITERIA

The choice of criteria against which an operational decision aid may be evaluated presents a unique set of problems to the aid evaluator. Criteria are standards against which the merits of an aid are judged. The ultimate criterion for the Navy would allow statements about how much an aid contributes to success during a combat situation. However, such a criterion is too remote for practical purposes. Backing away from this ultimate criterion means dealing with more and more intermediate criteria or proximal standards. Merit during a fleet exercise represents an intermediate criterion that (adequately?) approximates the ultimate. If a simulation or laboratory based evaluation is employed, a criterion is involved which is even more remote from the ultimate.

Within each category along this proximal-distal continuum, a number of considerations are apparent. The decision about a preferred criterion may be based, at least in part, on a number of attributes that an "acceptable" criterion must possess.

Aid Evaluative Criterion Attributes

Aid evaluative criteria must be acceptable, generalizable, reliable, objective, quantifiable, and analyzable.

Criterion Acceptability

In decision aid evaluation, criterion acceptability requires special attention because, much more so than the other criterion attributes, acceptability is a subjective matter. Acceptability refers to the degree that aid users will accept a criterion as an index of merit or relevance. Acceptability implies a value judgment--a reaction in terms of a subjective value assessment rather than in terms of objective characteristics. Accordingly, in evaluating the acceptability of a criterion, opinions about the criterion must be obtained from all stakeholders. Often, the criterion must reflect the divergent (and sometimes conflicting) opinions of these stakeholders.

This collectivistic approach rests on the assumption that defining or working out the details of the standards takes into consideration the views and needs of at least three important groups: (1) the developers of the aid, (2) its users, and (3) the evaluators.

Each group can make an important contribution to the criterion choice. Certainly, the goals or purposes of the developers of an aid are an essential ingredient in any aid evaluation. User's needs must be expressed because they are the latch-pins around which the aid must work. The evaluator will

also possess special needs which reflect evaluation cost constraints, time constraints, data collection constraints, and evaluation design considerations.

Critically evaluating a criterion's acceptability also provides a practical secondary benefit. It allows the criterion developer to reflect the various personal views of the interested parties. The results may be that each stakeholder, perceiving his individual contribution to the criterion choice, will be more able to identify with the criterion and will be more likely to accept findings based on it.

Generalizability

The importance of the generalizability construct can be appreciated in the context of statements by Feigl (1951), who called generalizations "empirical laws" and by Nagel (1967), who referred to generalizations as "experimental laws." This suggests the importance of a generalizable criterion to allow for "lawful" statements. Although Wolman (1973) argued that such statements are not laws but rather are simply inductive statements which draw strength on the basis of evidence, the importance of the generalizability construct to criterion choice does not seem to be disputable.

Criterion Reliability, Objectivity, Quantifiability, and Analyzability

The final features of criteria, reliability, objectivity, quantifiability, and analyzability are products of the logical-rational-empirical epistemologies basic to science.

Reliability refers to the complex property of a series of measurements that makes it possible to obtain similar results upon repetition of the measurements. A reliability statement provides information about the degree of repeatability of the measurements and how free they are from random variability.

Objectivity refers to the characteristics of a criterion which makes its measurement bias free and impersonal. As such, objectivity is related to quantifiability, which is concerned with the type of scale which the measurement yields. Because criterion data are generally reduced and treated by one or more statistical techniques, an underlying measurement scale is generally required which possesses ratio scaling properties.

An analyzable criterion is one which allows the determination of a variety of subscores as well as a total score. This characteristic provides the evaluator with insight into why an aid receives a given total score, i.e., the criterion aspects most and least affected by the decision aid.

Criterion Significance

In order to develop a criterion which possesses the several characteristics detailed above, the evaluator may find himself entrapped with a trivial criterion. Ultimately, adoption of a decision aid depends on operationally oriented statements, e.g., use of the aid results in a 20 percent decrease in the total cost of an operation, use of the aid results in a 50 percent increase in the number of enemy targets destroyed with a 7 percent decrease in own loss. Accordingly, the criterion should lead to or provide a basis for such operationally oriented statements.

Internal vs. External Criteria

By virtue of their design, many decision aids provide an internal criterion of user decision effectiveness. Often, aids involve an internal model which simulates a system and provides a recommended solution on the basis of a theoretic model, e.g., game theory, expected utility, or tradeoff analysis. When such a model is available, the user's decision may be compared with the preferred answers given by the internal model. A criterion based on such an approach, which is certainly tautological in nature, is called an internal criterion.

The use of such an internal criterion for evaluative purposes raises a number of interpretive problems. If the user makes decisions which agree with the internally generated recommendations of the aid, is the user satisfied with these decisions or has the user been seduced by the magic of the medium?

Operational decision aids are, after all, very sophisticated and virtually reek with the sparkle of modern technology and the authority of science. It seems possible that if a decision aid has these implicit attributes, then using the aid could create demand characteristics sufficiently strong to induce the user to accept, almost blindly, the aid's predictions.

External criteria are independent of the aid itself and represent a standard against which decisions made with and without the use of the aid can be compared. For example, in one study (Siegel & Madden, 1980), the criterion consisted of the decisions reached by a panel of experts about a set of scenario related problems. Within the evaluation process, the same problems and scenarios were employed, and the results were compared with the criterion data when an aid was used and when it was not used. However, if decisions made by experts do not agree with aided decisions made by others, is it proper to conclude that aided decisions are inadequate? Possibly, the aided decisions are superior to those made by the experts.

Locus of the Evaluation

Where the evaluation is conducted will also affect the choice of criterion. If the evaluation is conducted in a laboratory, a criterion can be selected which

possesses many of the psychometric characteristics described earlier. But, the results of a laboratory based evaluation may be less acceptable than one conducted in a more operationally oriented environment and in which the criterion is less psychometrically defensible. While the realism and fidelity of the evaluation are directly related to the acceptability of the results, they are inversely related to other criterion requisites. For example, one would be more willing to generalize from war exercise results to the combat situation than from a laboratory study to the combat situation.

Of course, it is possible to evaluate successively an aid along a progression which reflects realism/fidelity. One can perform an initial evaluation in a controlled laboratory study, and then, depending on results, proceed to a simulator based evaluation and finally perform an evaluation under operational conditions.

Other Criterion Problems

Keen (1975) suggests that post facto benefits analysis is not a suitable method for evaluating decision aids. Criteria derived after an aid is developed are considered to be unsuitable by Keen. He argues that an aid's purposes and how to assess their achievement should be a consideration before the actual development begins. The standards or criteria should be used throughout the development process to ensure that the aid meets the requirements. In addition, once an aid is sufficiently developed and ready for more rigorous testing, the criteria should be used to assess usefulness.

Criterion Choice and Multiple Criteria

The characteristics of an acceptable criterion do not subsume all considerations involved in the choice of a criterion. On the other hand, they do provide a basis for eliminating the more obviously unsuitable possibilities. The criterion choice will, in many ways, be dependent on the approach to the evaluation, the design of the evaluation study, and the goals of the evaluator. Is he interested in the relative efficiency of alternate aid designs? Is he interested in the extent to which the aid meets the developer's goals? The user's needs? Each of these will lead to the selection of different criteria. In prior work, the present authors assumed that usefulness is the dimension along which decision aids should be evaluated. The use of this attribute flowed from goals which sought to evaluate the potential of preliminary aids, as representative of classes of aids, rather than the absolute value of the aids in question.

In any event, the criterion choice will temper the "acceptability" of the results of the evaluation. Accordingly, the choice of criterion is like a game of Russian roulette. If the evaluator loses (selects the wrong criterion), the result will be disasterous to him. For this reason, multiple criteria are often

recommended. This approach not only allows the evaluator to hedge his bets, but also allows him a fall back position. Finally, it allows a measure of convergent validity. Two different measures which mutually support each other are less vulnerable than either method alone.

Prologue to Subsequent Chapters

In the next two chapters, three different general categories of methods for aid evaluation are described and their various strengths and weaknesses are probed. Chapter III discusses analytic methods, while Chapter IV presents experimental/quasiexperimental methods. Within each chapter, a number of techniques is considered. Each technique is elaborated to demonstrate how it is applicable to decision aid evaluation. The strengths and weaknesses of the various techniques are discussed in context.

CHAPTER III

ANALYTIC METHODS FOR DECISION AND EVALUATION

Analytic methods for decision aid evaluative purposes constitute a set of techniques which are largely paper-and-pencil oriented, attempt to describe traits or characteristics by following reductionistic methods, and tend to represent ways to assess selected attributes of an aid. One common assumption of such techniques is that an aid can be evaluated in terms of its elements and that the benefits of an aid can be understood by studying the elements. Some of the techniques also assume that the verbal behavior of individuals can be used to form a reliable index of an aid's attributes. The techniques are generally structured to sample specific data about the attributes of an aid.

Cost/Benefit, Cost Effectiveness, and Cost/Utility

Cost/benefit, cost/effectiveness, and cost/utility analyses attempt to answer the question, "Is it worth the cost?"

Levin (1975) distinguished among cost/benefit, cost/effectiveness, and cost/utility as follows:

> While cost/benefit analysis enables a direct comparison of costs and benefits stated in monetary terms and cost/effectiveness represents an attempt to evaluate directly the costs of alternative ways of achieving particular outcomes, cost/utility analysis incorporates the decision maker's subjective views in valuing the outcomes of alternative strategies." (p.94)

The merit of each of these techniques is dependent on the integrity of each numerator and denominator in the respective ratio. Cost forms the numerator, the denominator, or both in all three of the ratios. Monetary information is often difficult to acquire and monetary estimates are notoriously faulty. Accordingly, the attractiveness of these cost oriented approaches may be misleading.

It is also clear that each of the denominators represents a different frame of reference. The cost/effectiveness approach seems most appropriate for evaluating costs of reaching a given decision of given quality (an objective specified in advance) with and without reliance on a given decision aid. However, a complementary interpretation of cost/effectiveness can consider constant costs and

examine increments or decrements in decision effectiveness as a function of alternative means. The cost/benefit approach seems most appropriate when benefits can be accurately stated in monetary terms and cost/utility analysis seems most appropriate for assessing the user's appraisal of decision aids with varying features and of varying degrees of sophistication.

While the three approaches are by no means mutually exclusive, each seems appropriate for some aspect of operational decision aid evaluation and combinations, especially in sequence, seem particularly meaningful. For example, cost/benefit assessment subsequent to establishing basic cost/effectiveness seems advisable. In turn, cost/utility assessment, qua acceptability, may be an essential prerequisite for both.

Cost Determination

A decision aid considered for actual and significant use will, in most cases, be computer based. In effect, it will be a type of decision model (software) implemented largely within a general purpose environment (hardware). In some cases, special hardware (e.g., display, communication) might be involved. To the extent that these special hardware system components are not off-the-shelf adjuncts, the associated costs will be for:

- design
- construction
- maintenance

The aggregate costs for these purposes can be considered separately, as an increment to overall hardware costs.

The cost categories of major interest will be those associated with software, i.e., designing and programming the decision aiding model. Siegel and Wolf (1981) suggested the following relationships for costing any sort of computer model:

Life Cycle Cost

$$LCC = MDC + MTC + MMC + \sum_{n} MUC_n$$

Model Development Cost

$$MDC = MCC + MPC + MDAC + MDDC$$

Model Test Cost

$$MTC = MSTC + MVTC$$

Model Maintenance Cost

$$MMC = MEC + MECC + MMDC$$

Model Utilization Cost

$$MUC = RSC + \sum_{j} (CMC_j \cdot CTR_j)_u$$

Model Conceptualization Cost

$$MCC = \left(\sum_i DLH_i \cdot DLR_i\right)_c + MTO_c$$

Model Programming Cost

$$MPC = \left(\sum_i PLH_i \cdot PLR_i\right)_p + \sum_j CMC_j \cdot CTR_j + MTO_p$$

Model Sensitivity Test Cost

$$MSTC = \sum_i \left(TLH_i \cdot DLR_i\right)_s + \left(CMC_j \cdot CTR_j\right)_s + MTO_s$$

Model Validation Test Cost

$$MVTC = \sum_i \left(TLH_i \cdot DLR_i\right)_v + \left(CMC_j \cdot CTR_j\right)_s + MTO_s$$

where:

MDC	= Model Development Cost
MTC	= Model Test Cost
MMC	= Model Maintenance Cost
n	= Number of Model Applications or Uses
MUC_n	= Model Use Cost
MCC	= Model Conceptualization (problem definition) Cost
MPC	= Model Programming Cost
MDAC	= Model Data Acquisition Cost
MDDC	= Model Design Documentation Reporting Cost
MSTC	= Model Sensitivity Test Cost
MVTC	= Model Validation Test Cost
MEC	= Model Enhancement Costs
MECC	= Model Error Correction Costs
MMDC	= Model Maintenance Documentation Costs
c	= Conceptualization Phase
p	= Programming Phase
s	= Sensitivity Test Phase
v	= Validation Test Phase
u	= Utilization Phase
i	= Personnel Types Involved
DLH	= Design Labor Hours
DLR	= Design Labor Rate
MTO	= Material, Travel and Other Costs

PLH	= Programming Labor Hours
PLR	= Programming Labor Rate
CMC_j	= Cost of Computer Related Elements Per Unit Time (computer, memory, line, costs)
j	= Types of Computer Related Costs
CTR_j	= Time Required of Computer Related Elements
TLH	= Test Labor Hours
TLR	= Test Labor Rates
RSC	= Run Setup Cost

Tradeoffs

When cost figures are available, various tradeoffs may be completed.

In the general case, possibly the greatest advantage of decision aids lies in their ability to provide the basis for economical increase in man-machine equipment system performance. Figure 1 presents a hypothetical representation of the gain in system effectiveness when various amounts of money are invested either in physical design improvement or in decision aid implementation for the operator of the system. In the illustrative example, if less than about 300 cost units are available, the system manager would be well advised to invest most, if not all, of his available funds in a decision aid development rather than physical equipment redesign. In the 450 cost unit area, physical design improvement produces an increase in system effectiveness that equals the gain to be anticipated from a decision aid.

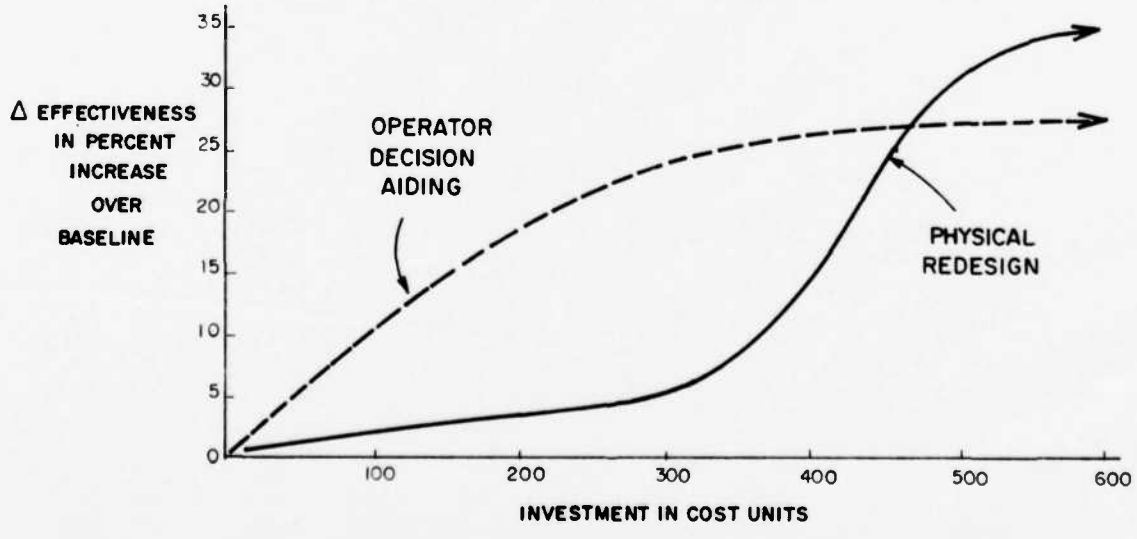


Figure 1. Hypothetical example of tradeoff between decision aiding and physical equipment redesign.

When more than 450 cost units are available for improvement, these data show the best investment to be in a physical redesign. Note that the asymptotic situation is achieved for both alternatives in the range of 500-600 cost units. Exemplary of the type of tradeoff are the not-so-obvious results generated from the Figure 1 data. These are shown in Table 1. For each level of funding available for system improvement, Table 1 shows the way to greatest payoff:

0-300 cost units - buy decision aids

450 cost units - buy training or physical redesign

500 cost units - buy physical redesign

Table 1 also shows that no advantage can be gained by applying more than 1200 cost units and that the largest payoff in terms of Δ effectiveness per unit funding applied is in the 200 cost unit range (see* Table 1).

Table 1

Value of Δ Effectiveness Resulting from the Best Choice of Funds Application for Given Amount of Funds Availability (Hypothetical Example)

If this no. of cost units were the available funding for improvements	and these funds were divided among: Physical Redesign	Decision Aids	the resulting Δ effectiveness would be:	The ratio Δ effectiveness/cost unit is then:
100	0	100	8	.080
200	0	200	18	*.090
300	0	300	23	.077
400	0	400	26	.065
450	0	450	27	.060
450	450	0	27	.060
500	500	0	32	.064
600	500	100	40	.067
700	500	200	50	.071
800	500	300	55	.069
900	500	400	58	.064
900	600	300	58	.064
1000	600	400	61	.061
1100	600	500	62	.062
1200	600	600	63	.063

User Reaction

A variety of techniques is available for obtaining information about the specific attributes and the general utility of an aid. Each depends on som familiarity with the aid and most yield a quantitative score. The interpretation of these scores is often left to the evaluator. Norms are usually not available. Accordingly, the techniques seem most applicable to the comparative evaluation of alternate aid designs or of alternate aids. Nevertheless, application of one or more of the techniques to an individual aid should yield considerable insight about the attributes and the deficiences of that aid.

Quite obviously, it is not possible to include here all such techniques or even a full description of the selected techniques. Nevertheless a variety of such techniques is presented below in order to provide insight into the range of techniques available. Each included technique was selected because it is somewhat unique. Original sources are cited within the discussion of each technique for the benefit of the evaluator who wishes a fuller description of the technique described.

Polydiagnostic Method

The Polydiagnostic Method was developed by Bennett (1956) as a multivariate social and clinical research tool. It is included here because the technique possesses a degree of novelty. The method seems appropriate for assessing reactions to the design features of an aid.

Basis for Method

According to Bennett, a number of subjective forces influence a user's reaction to a product. Product design is assumed to be assessible through a number of levels of user acceptance. Bennett developed three concepts for analyzing acceptance: (1) the "personality" of the design, (2) the user's personal reactions to the product, and (3) the reactions that users predict for other possible users of the product.

The concept that design products are perceived to have "personalities" appears to be based on common sense. Just as people can describe their impressions of the nature of friends, they can describe this impression of "an aircraft, a suit of clothing, or a typewriter" (Bennett, Kemler et al., 1958).

The second concept in acceptance is centered on the personal reactions or feelings of users relative to the product. Are users satisfied or dissatisfied with the product? What are the high points of their reactions? Does the design engender positive or negative attitudes about the product? Knowledge of these feelings can direct designers to the product's assets and liabilities.

The third concept is concerned with how a user perceives the reactions of other users toward the product. This is taken to indicate "...what do the users see as the general acceptability of a design product, aside from their own particular feelings?"

If, as assumed by the Polydiagnostic Method, the acceptance or rejection and success or failure in a design effort is a function of subjective feeling and thinking, then some means of measuring these socio-psychological experiences is required. The essence of the method is that in order to understand how the users feel and think about a product, they must make judgments. The important point of the method is to make the judgmental task objective, structured, and quantitative. This increases the probability of achieving an acceptable evaluation.

Bennett recognizes the inherent need for precise data on "feelings and impressions" and indicates that the most effective techniques for "obtaining objective information relevant to subjective experience" are the psychophysical methods.

Application of the Method

The Polydiagnostic Method is a special case of the method of multiple forced-choice rankings. It combines principles developed from the psychophysical rating, ranking, and choice methods. The steps in the application of the method, as described by Bennett, Kemler et al., (1958), are:

1. A number, n, of qualities important to the specified problem is selected to function as a set. An unlimited number of n quality sets can be selected.

2. A number, m, of items to be studied (decision aids) is defined.

3. A number, h, is selected which consists of steps, degrees, or points on the proposed rating scales. The number, n, must be divisible by h.

4. Each set of n qualities is presented to a user.

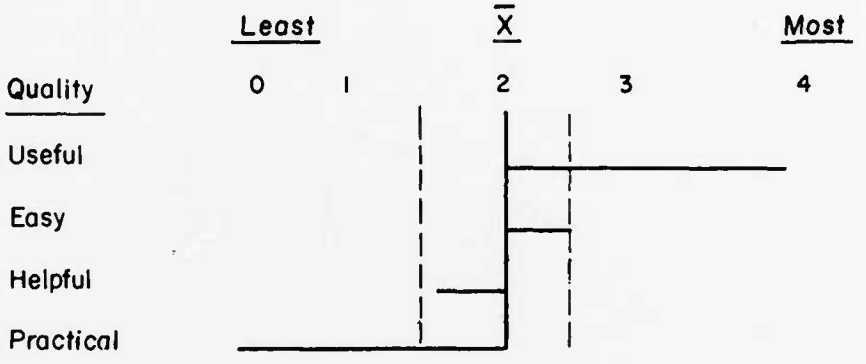
5. The user is asked to select q of the n qualities in each set according to some instruction, e.g., the q qualities that best describe the product. The value q is determined as equal to n, divided by the number, h, of rating scale categories.

6. From the n-q remaining qualities, q more are chosen following the same instruction.

7. This process of choosing q qualities in each set continues until q qualities remain.

8. Each of the qualities chosen in the first set of q are assigned an ordinal score of q-1. The next q qualities are scored q-2, and so forth until the final unchosen qualities are scored q-q or zero.

Group means are calculated by quality and displayed for convenient comparison as shown in the example below:



In this example, a five point scale is used, and 2 is the expected value of each rating if chance factors alone operate. The dashed lines represent boundaries above or below which scores are significantly (level of confidence = .05) different than the chance mean.

Advantages and Disadvantages of Method

There are a number of advantages and some disadvantages in utilizing the Polydiagnostic Method as a tool for evaluating decision aids. The major advantages of the method include its objectivity and flexibility, along with its ability to quantify very complex judgments in a direct manner.

The disadvantages of the Polydiagnostic Method relate to the type of data it yields, the problem involved in selecting qualities, and the choice of the participants. Although collected by sophisticated techniques, the data yielded do not allow for a statement of causal relationships. For example, one cannot know, except by extrapolation, that "acceptance" is caused by quality "X!" One observes an association but causality must be presumed.

Selecting qualities to be judged presents a problem because the selection may introduce evaluator bias. The result may be a neglect of important qualities. The choice of qualities, accordingly, needs to be fully verified for completeness in consultation with all stakeholders in the evaluation. Once a list of qualities is verified and found to be thorough, application of the technique becomes almost mechanical.

The final area of concern relative to the disadvantages of the Polydiagnostic Method is the effects of the users in the evaluation. In decision aid evaluation studies, it is not necessary that ultimate users serve as participants so long as those that participate have broad operational and practical experience in the area which the aid supports. However, direct analysis in terms of current practice requires the participation of currently fluent users. Only those using the current standard operating procedures can be expected to make reasonable assessments of an aid vis-a-vis these procedures. This type of participant experience also seems necessary for acquiring information about the characteristics of the "ideal" decision aid for a given purpose. Otherwise, a misleading set of data may be obtained.

Application Examples

It becomes evident that the Polydiagnostic Method may be employed to provide a subjectively based assessment of an aid's strong and weak qualities. In addition, the method can be used for comparative evaluations. Here, the same replicating conditions, questions, and qualities would be used but for two or more aids or aid designs. In the same vein, Polydiagnostic Method may be employed to compare an existing decision aid with the user's "ideal" conception of the aid.

The prior uses of the method are concerned with analyzing the overall "personality" of an aid. A modified use of the method is also possible. A dimension or quality can be selected and that quality can be evaluated. For example, display adequacy is one quality of a decision aiding system. The displays may be assessed along any quality dimension which the evaluator believes to be important, e.g., utility, readability, adequacy of content, completeness, etc.

The Polydiagnostic Method can be used to evaluate an aid in isolation, comparatively, or in some combination of both approaches. For example, assume that three options are under consideration: (1) current, unaided standard operating procedures, (2) current procedures with an available aid, and (3) current procedures supplemented by an "ideal" aid. The Polydiagnostic Method could provide important information relative to the choice of option.

Overall Summary and Evaluation

It is apparent that the Polydiagnostic Method offers a flexible approach to decision aid evaluation when subjective opinions and reactions are sought.

The method's flexibility enables evaluation of an aid or comparative evaluation from a diversity of perspectives. A number of qualities may be assessed by application of the technique. While the choice of qualities may be subject to evaluator bias, the various Bennett studies (e.g., Allen, Bennett et al., 1959) suggest that such problems are not disqualifying.

There is little reason to believe that the reliability and the validity of the data yielded by the technique will not be equal to, if not greater than, that yielded by other subjective/interview methods. However, reliability investigation should probably be included within any individual use of the method. Validity will vary from use to use and with the user's intelligent choice of qualities for investigation.

Analytic Profile System

The Analytic Profile System (Siegel, Fischl et al., 1975) is an objective, paper-and-pencil technique for deriving an indication of the human factors adequacy of visual displays. It is included here as an example of a structured questionnaire. The system is derived from a multidimensional scaling analysis of the interface between a display and its observer. The multidimensional scaling analysis indicated the display-observer interface to be describable in terms of seven unique dimensions: (1) stimulus numerosity, (2) primary coding, (3) contextual discrimination, (4) structure scanning, (5) critical relationships, (6) cue integration, and (7) cognitive processing activity. Descriptive items (a brief prose statement about an aspect of a display) were written which reflect the meaning of the various display dimensions. A large number of items was developed and scaled. From these, final items were selected on the basis of reliability and homogeneity. Then, a final instrument was assembled in forced choice format and validated: concurrent validity, $r = .75$; predictive validity, phi = .87. Accordingly, the Analytic Profile System appears to have been rather carefully developed and validated. It makes satisfactory use of the forced choice and the factor analytic techniques and seems to possess acceptable psychometric characteristics.

Advantages and Disadvantages of the Analytic Profile System

The care devoted to the construction of the system (Fischl & Siegel, 1970; Siegel, Fischl et al., 1975) supports a contention favoring confidence in the indications of its application. The system does not possess the flexibility identified with the Polydiagnostic Method. However, the Analytic Profile System permits a complete examination of an aid's displays relative to each of the seven information transfer factors it considers. The system permits an independent analysis along the seven dimensions. The result is a score for each dimension that is directly interpretable in terms of the display-observer interface adequacy, e.g., a low score on the Cue Integration dimension suggests that the

aid's displays do not facilitate the function of information integration. Accordingly, the Analytic Profile System provides a direct indication of areas of needed improvement. In addition to the scores corresponding to the separate dimensions, a score is made available which represents the overall adequacy of the displays under consideration. This score can be useful in a variety of comparative and relative analyses.

The Analytic Profile System is easy to apply. If the technique is used to assess existing displays, then no preparation is necessary. However, if the evaluator's purpose is to compare a present display with some nonextant alternative, then consideration must be given to the design of the alternative display.

The drawback of the Analytic Profile System appears to be that it does not provide direct information about how a display might be improved. Each dimension score indicates the degree to which a display reflects the requirements of that dimension. However, given a low score, the technique does not state how to improve the display. This insight is left for the display developer to provide.

There may be aspects of the display-observer interface which are not evaluated by the Analytic Profile System. Accordingly, it does not seem that the Analytic Profile System can be employed as a sole evaluative technique. Rather, it seems that it represents a useful accessory to other evaluative techniques and, as such, provides information that is not provided by these other techniques.

Moreover, the technique is only applicable to display-observer interface analysis. Other aid aspects cannot be analyzed by this technique.

Evaluation Applications

Given the dimension scores and the total score yielded by the Analytic Profile System, several decision aid applications can be given.

The Analytic Profile System can be employed to evaluate a display from the absolute point of view and to answer questions such as: How adequate is the display? On what dimension is it weak? On which is it strong? Answers to such questions provide insight into areas for needed display improvement.

The potential of the Analytic Profile System for comparative evaluation is also apparent. Two or more alternative displays can be objectively assessed for their relative merit through Analytic Profile System application. Similarly, the Analytic Profile Systems can be employed to assess the improvement brought about by modification of an individual display. Here, values taken before and after display modification would be compared.

Display Evaluation Index

The Display Evaluation Index (Siegel, Miehle et al., 1964) was developed as a technique for evaluating a system's information transfer capability. Its goal is to assess the ability of displays to transfer information to operators and of the operator to process the information and perform the required and appropriate control action(s). The index provided by the Display Evaluation Index possesses both concept and empirical validity. The index is based on a number of information transfer principles. Each principle has been shown to possess construct validity (Siegel, Miehle et al., 1964). The principles incorporated within the Display Evaluation Index are: other things being equal, that system is best which:

1. requires the least operator information processing per subtask unit
2. has the greatest directness between information transmitters (displays) and receivers (controls)
3. has the least difference between the amount of information presented by an indicator and that required by a control action
4. provides for redundancy of information
5. requires the least intermediate data processing by the operator before he can perform the required control action
6. has the least number of information sources and sinks
7. imposes the least amount of time stress on the operator as he performs the information processing
8. has the least number of transfers which cannot be accomplished within a prescribed time
9. possesses the least number of critical transfers
10. has displays and controls that are optimally coded.

The Display Evaluation Index is based on the assumption that system design aspects, as reflected in the principles, affect the operator's ability to perceive displayed information, process it, and take the appropriate action. These form the basis for deriving scores called complexity, directness, data transfer,

encoding, time, match, and critical links. Each score is an expression of one or more of the principles, e.g., the time score reflects principles 7 and 8. The scoring is accomplished through well-defined procedures using quantitative formula. The resultant scores are entered into a formula which produces the final index. Accordingly, the method makes available seven subscores, one for each factor and a total score--the index.

Use of the Display Evaluative Index

The method can be used to investigate the information transfer characteristics of an aid in relationship to each factor. Additionally, the total score provides an overall index of merit. Those factors that are well integrated within a display can be ascertained as well as those that detract from information transfer. This means that questions can be answered about why an aid's display and display-control relationship need improvement. This capability derives from the relation between each subscore and one or more of the information transfer principles.

In a decision aid, there are often any number of displays, each different from the others. An evaluator might want to know how effectively each display meets the information transfer needs. The Display Evaluation Index technique seems appropriate for this purpose.

Another use of the Display Evaluation Index technique in decision aid evaluation is comparative evaluation of alternate displays and the associated actions. Since the technique considers both displays and the associated actions, it would seem particularly applicable to menue evaluation, error displays, and cueing displays as included in many decision aids.

Advantages and Disadvantages of the Display Evaluation Index

As for the two prior techniques, the Display Evaluation Index method is standardized both in terms of method of application and method of scoring. It is objective and its construct validity derives from the principles on which it is based. The empirical validity of the method was investigated in a number of studies (Siegel & Federman, 1967, Siegel, Miehle et al., 1963) and found to be adequately high. Similarly, the interanalyst agreement for moderately well - trained analysts was reported to be quite acceptable (Siegel, Miehle et al., 1964). The technique is applicable while an aid's displays are in the conceptual stage of development. Hence, the method can be employed as a diagnostic tool for individual display improvement or for comparative evaluation of alternate displays early in the decision aid development cycle.

Application of the Display Evaluation Index technique requires composition of a "transfer chart." This chart is a symbolic representation of the information transfer involved in aid use and constitutes a basic tool for using the method. The chart symbolically represents the displays, the controls, the links between the two, and the information processing involved within each link. Bias may be injected into the method: (1) in the development of the transfer chart, and (2) in the selection of analysts. In spite of the quite acceptable interanalyst agreement reported by the method developers (Siegel, Miehle et al., 1964), it remains possible that different analysts might not always produce the same end products. In aid evaluation, this issue may best be addressed by using a reasonable number of analysts who are reasonably independent from the aid's primary developer. This procedure would allow the development of an index of interanalyst agreement befor employing the various scores for aid evaluative purposes.

Finally, while the Display Evaluation Index technique yields a total score and a set of subscores, translation of the scores into display-observer interface improvement is left to the analyst. Hence, the technique is diagnostic but not prescriptive. While subjective diagnosis represents an unquestionable advantage, it is not a sufficient end in itself. As for the prior two techniques, the Display Evaluation Index cannot be judged to be a stand alone technique for decision aid evaluation.

Shalit Perceptual Organization and Reduction Questionnaire

The Shalit Perceptual Organization and Reduction Questionnaire (Shalit, 1978, 1979) yields a quantitative assessment of the degree of perceptual organization and a qualitative assessment of the structure of certain defined aspects of a person's environment. While not designed for and not previously employed in person-equipment interface evaluation, the technique has a number of attributes which make it interesting in this context. The Shalit technique is designed to analyze dimensions of the "perceptual structure of a specific universe" (Shalit, 1979). In the context of the present work, the decision aiding device and the operational context would constitute the "environment."

Shalit (1978) attempted to add to Kelly's (1955) concepts by considering the difficulty involved in coping with a variety of situations. The situations involve different demands which are described in terms of three dimensions: differentiation, articulation, and loading. Differentiation relates to the number of factors perceived in a situation; articulation is the range and clarity with which the factors are ranked, and loading is the positive or negative emotional tone associated with a situation. Shalit (1978) found that the articulation and loading characterizing a (decision aiding) environment predict the ability to cope with the environment. Thus, "the less structured and organised (sic) the environment was--in terms of lack of articulation and higher loading--the more difficult to handle" (Shalit, 1978).

The principal suggestion is that those decision aids are superior which articulate the environment and which have limited emotional connotation. Articulation, organization, and structuring are said by Shalit to be fundamental to proper understanding of the important aspects of a situation and how it can be successfully dealt with.

The functions of a decision aiding device are often to help a user to understand properly what is important and to organize and structure relevant information so as to facilitate the making of appropriate decisions. Therefore, a useful decision aid should increase a user's ability to differentiate and articulate the operational environment, while at the same time it should minimize the emotional loading so that the environment can be better coped with.

Description of the Shalit Perceptual Organization and Reduction Questionnaire

To employ the Shalit technique, the user lists all factors which he considers relevant in the context under consideration.

Because Shalit believes that writing the factors in the form of a list might affect the analyst's judgment, the factors are entered on a wheel shaped response form, as illustrated in Figure 2. After entering the factors, the analyst ranks them according to their importance to him, with the understanding that ties can occur. The assigned ranks are marked on the innermost circle of the wheel (Figure 2).

Following the ranking, each listed factor is rated positive or negative (pleasing or displeasing) on a five point scale. The degree and nature is indicated by placing ++, +, 0, -, -- in the outer compartment of each factor.

The information obtained can be conceptualized as a "structured" picture containing the number of factors, the clarity and range of the ranks, their loading, as well as information about the content of the universe.

Scoring is straightforward, with each step reasonably well documented. Scoring falls into two categories: (1) quantitative or scoring for structure and (2) qualitative or scoring for content. The quantitative scoring involves calculating three raw scores: (1) a differentiation score representing the number of factors listed, (2) an articulation score based on the range of rankings, and (3) a loading score based on a count off the + and - signs entered.

From those three, several additional indices are constructed: (1) a reduction score, (2) an emotionality score, and (3) an intensity score. The reduction score reflects cognitive organization. It varies inversely with the degree with which factors have been ordered into ranks. Situations with lower scores have a higher cognitive organization and can be expected to yield a better ability to cope with the demands of the environment.

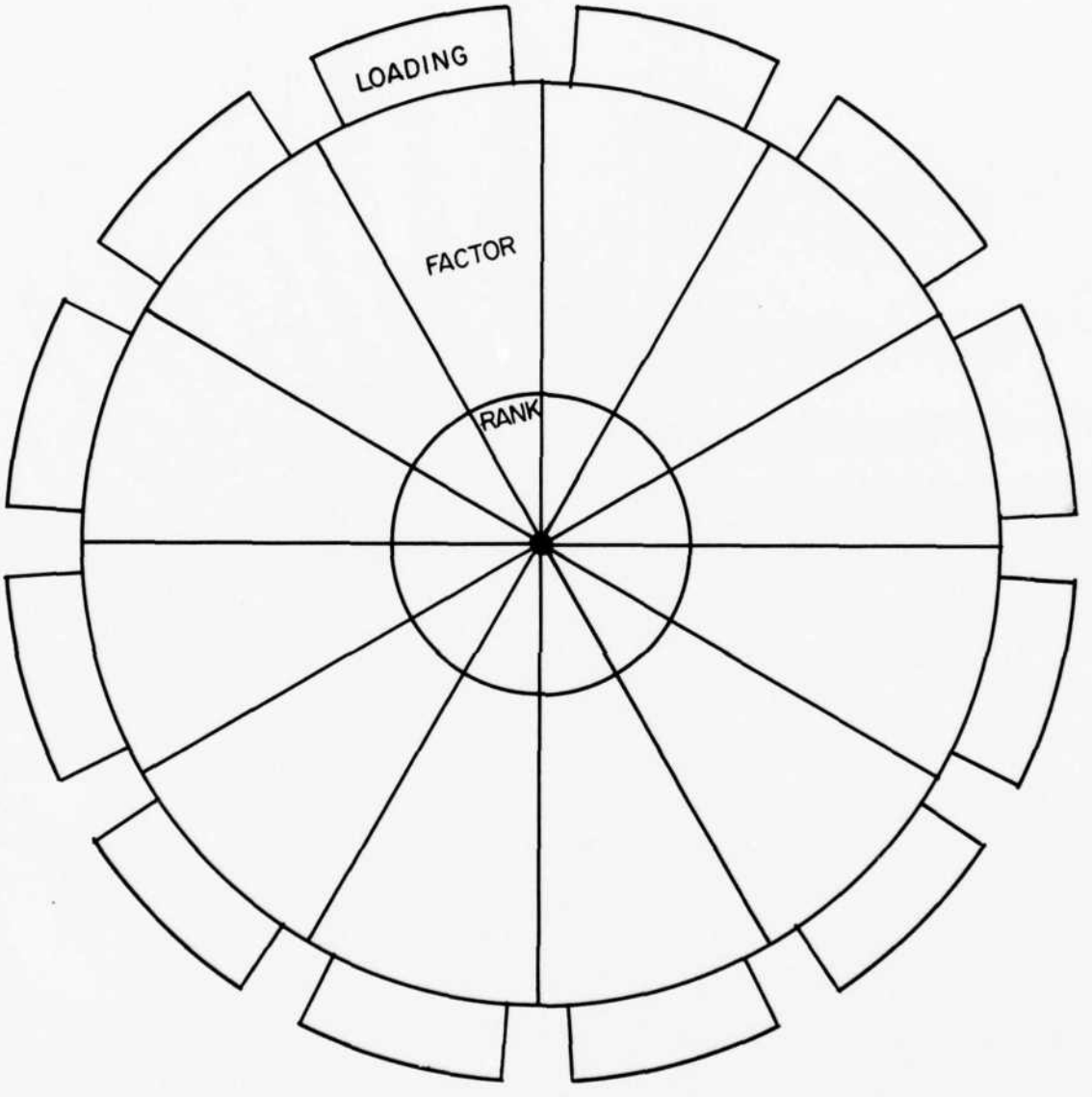


Figure 2. Shalit response form.

The emotionality index reflects the total emotional loading associated with a situation. The higher the index, the higher the emotional potential associated with the perceived universe. The intensity index reflects the interactions of emotional loading and cognitive organization. It suggests both the relative strength and direction of the emotion associated with each factor.

The scoring for content, or quality, uses a sentence mapping approach. Each of the factors listed in the wheel is first scored on four dimensions: (1) aspect, (2) target, (3) mode, and (4) valence. Aspect refers to the characteristics of the situation and can be classified as either cognitive, affective, instrumental, or physical. The target dimension describes the parts of the environment and does not seem to be related to decision aid evaluation. Mode suggests the kind of relationship that exists between the user and the situation. Mode can be either self focused, interacting, or other focused. Valence expresses the associated feelings as overtly expressed in the +, 0, or - signs.

Potential Application in Decision Aid Evaluation

On the basis of these concepts, it seems that the Shalit technique can be used as a tool for evaluating decision aids. The reduction, emotionality, and intensity indices, along with the profile compiled from the qualitative scores, seem applicable to evaluations in a number of different ways. The mode and structure of perceptions of an operational universe in which the decision aid is included can be obtained. The same information can be obtained from an operational universe in which the decision aid is excluded. The two information sets might then be compared. Possible changes in perceptive mode and structure as a function of decision aid effects might be analyzed and compared across the two situations.

Other perceptions might also be analyzed and compared with the aided situation, e.g., a user's ideal aided operational situation or how other groups (experts, peers) might perceive the various situations.

Other evaluative questions might also be answered through application of the Shalit technique. Does the use of an aid enable a user to make a more appropriate differentiation of the operational environment in which his task must be accomplished? Can he better assess the importance of each factor that must be considered to arrive at reasonable operational decisions? Does using the aid decrease the negative connotation in the user's perception of the operational environment? Does it increase the positive effect? Does the decision aid supply data congruent with a user's mode of perception? If not, how can the aid be modified so that its data and the user's perception are more compatible? Is the degree of coherence in the organization of the perceptual data facilitated by using the aid? Do temporal factors affect the indices and profiles? Does the environment containing the aid compare favorably with an environment containing an ideal aid?

Advantages and Disadvantages of the Technique

The Shalit technique seems relatively easy to apply and score. It seems to be reasonably standardized and to possess content validity relative to decision aid evaluation. Its empirical validity for decision aid evaluation has not been established. The reliability of the method was reported by Shalit (1978) to have been investigated in a number of populations and found to be acceptable.

The decision aid evaluative information that might be provided by the Shalit technique is of a character which is different from that which is provided by the previously discussed analytic evaluative tools. Accordingly, the technique might best be employed as a supplemental tool in association with the use of some of the prior analytic methods.

Multiattribute Utility Analysis

Multiattribute Utility Analysis is a scaling method developed by Edwards (1971) for evaluating the utility of various courses of action in terms of "value" or "benefit." In the present context, the technique may be employed to determine the utility of an aid or to evaluate comparatively two or more aids in terms of their utility. The technique represents a subset of evaluation methodologies classified by House (1980) as the "decision making approach." In developing the method, Edwards (1971) dismissed as a myth the concept of a decision maker who makes decisions which, on the average, maximize his values. Edwards also avoided the distinction between risky and riskless choices and ignored probabilities. He treated all utilities as ordinal values, disregarded interactions, and presented a 10 step algorithm for utility assessment.

Application of Multiattribute Utility Analysis

The first three steps of Edwards' procedure require establishment of the important variables around which the utility analysis is developed. This demands that the goals of the item (decision aid) whose utilities are to be established be identified.

In the case of a decision aid, these would be provided by the developers of the aid or by operational personnel. Next, the goals are ranked in importance and weighted so that the sum of the goal weights is equal to 100. Then, the extent of achievement of each goal is estimated. The estimate is performed on a 0 to 100 scale. The magnitude assigned on goal "X" represents utility for that goal.

Finally, an overall utility is calculated by the equation:

$$U_i = \Sigma_j w_j u_{ij}$$

where $\Sigma_j w_j = 100$

U_i = aggregate utility

w_j = normalized importance weight of dimension j

u_{ij} = utility of i in dimension j

Prior Uses of Multiattribute Utility Analysis in Decision Aid Evaluation

Multiattribute Utility Analysis has been employed within two decision aid evaluative studies (Siegel & Madden, 1980; Madden & Siegel, 1980). The results of its application provided a number of insights vis-a-vis aid improvement, aid goal achievement, and perceived utility. The findings were apparently reliable and were not challenged. This is somewhat surprising given that Edwards never validated the method. House (1980) noted this gap and suggested that any support to the multiattribute utility analytic technique must rest solely on the virtue of its utility.

Interviews

Interviews (unstructured, semistructured, and fully structured) represent the final analytic technique to be discussed here. Such interviews represent a method for acquiring a wealth of subjective information about the acceptability, sensitivity, design attributes, and utility of any decision aid.

Interviews may be specifically tailored for the aid in question and for target interviewees. Accordingly, for a given aid, there may be specific interviews designed for inquiry into the reactions of operational personnel, design personnel, human factors analysts, computer personnel, and the like. Moreover, interviews may be implemented at one or more stages of the aid developmental process.

As an interview becomes less structured, it allows for an increasing amount of probing by the interviewer, but the reliability of the information probably decreases.

Interviews have been successfully employed in a number of prior decision aid evaluations (Siegel & Madden, 1980; Madden & Siegel, 1980). In this work, interviews were employed as a supplement to a more formal evaluation, and the interview results were employed to interpret and elaborate the quantitative findings.

Types of questions which may be investigated through interview methods include, but are not limited to:

1. The adequacy (and inadequacy) of a decision aid for each of its goals/objectives and how to improve the aid's adequacy
2. Needed areas of improvement and why improvement is needed
3. Specific characteristics of the aid most attributable to its projected success
4. Specific characteristics that are likely to present problems to the users and how
5. New insights that the aid provides to its user; if none are provided how might the aid do so
6. Pertinence of information; what information could be eliminated
7. Information that the aid should provide, but does not
8. Comprehensibility of displays
9. Adequacy of error messages
10. Ease of use, with examples of difficult areas and suggestions for improvement
11. Adequacy of data updating procedures as conditions change
12. Adequacy of the input formats with suggestions for improvement
13. Other desirable features to be included in the decision aid and features which might be excluded
14. Advantages and disadvantages of the aided as compared with an unaided procedure.

Summary Review of Analytic Methods

The examination of analytic methods was not intended to be exhaustive. Rather, the intent was to provide a broad perspective of the range of analytic techniques which are available for aid evaluation.

Measurement Considerations

Measurement may be conceived as the classification of observations into categories according to specific rules (Cliff, 1973). The analytic methods are concerned with how best to make the necessary observations and classify them. The similarities among the various methods in regard to observation and classification can be appreciated in terms of Guilford's (1954) "internal response continuum" or what others refer to as "intervening variables." Guilford (1954) distinguished three continua which are relevant to the analytic methods. The first two, the stimulus and the response (or judgmental) continua, are observable, while the third, the internal response continuum, is not. By and large, each analytic method is based on one or more intervening variables (internal response continuum) to account for observations, judgments, or responses in relationship to the stimulus (aid feature) continuum. The intervening variables among the various methods are all essentially different processes which demand different response patterns. In addition, they are aimed at assessing different stimulus characteristics.

All of the analytic techniques relieve the evaluator from the criterion selection problem because individual criteria are embedded within the various techniques. Unfortunately, only one, the Polydiagnostic Method, also supplies a significance level as a byproduct of routine use. However, significance bands can be established for the information provided by most of the other techniques.

Unfortunately, the embedding of criteria within the analytic techniques may limit their utility. The most general statement that one would want to make after an evaluation is that the decision aid is useful for fleet operations. The information made available by the analytic techniques does not permit such a general statement.

Application Generality

Most of the analytical methods are specific in focus, but they appear to have applicability to most types of decision aiding evaluative exercises.

The analytic techniques are not location bound, and each technique can be used alone or with other analytic techniques.

Each analytic technique can be incorporated within large scale, controlled or quasiexperimental designs or in observational situations. In addition, some of the information yielded by the methods can be used to develop causal or structural equation models.

CHAPTER IV

THE EXPERIMENTAL METHODS, CONTROLLED AND QUASI

Empirical research is generally divided into three categories: (1) experimental, (2) correlational, and (3) observational. The experimental approach, characterized by rigorous control and direct manipulation of independent variables, is the essence of empirical science. The hallmark of correlational research is association, and the observational approach emphasizes description. Associated with each category is a difference in where the research is usually conducted--from laboratory to field (fleet). The laboratory to field vector might be conceived as increasing in realism and decreasing in precision. This continuum might also characterize the differences between the aid evaluations possible in a laboratory, in a more realistic simulated environment, and in the fleet.

Concomitant with these differences is a further suggestion that aid evaluations conducted in a simulated environment will be less experimental and more correlational or observational in nature. A number of research characteristics also seem to vary as decision aid evaluation is conducted along the laboratory to fleet continuum. Table 2 lists a number of these characteristics and assesses them relative to the location of the evaluation. These assessments suggest that advantages and disadvantages exist in each type of situation. Accordingly, depending on the locus of any evaluation, different evaluative statements can be made about a decision aid.

Table 2

Research Characteristics of Various Situations

Research Characteristics	*Laboratory*	*High Fidelity Simulation*	*Fleet*
Validity	Low	Moderate	High
Reliability	High	Moderate	Moderate
Precision of Dependent Measure	High	Moderate	Low
Objectivity	High	Moderate	Moderate
Control of Independent and Confounding Variables	High	Moderate	Low
Design	Experimental	Correlational	Observational
Generality	Low	Moderate	High
Statistical Tractability	High	Moderate	Low
Realism/Fidelity	Low	Moderate	High
Outcome Statements	Causal	Associational	Descriptive

Control

The ability to exercise research control is probably the characteristic that contributes most to the differences across locations. Control over independent, exogenous, and endogenous variables engenders stability, reliability, and interpretability to date. To control is to eliminate or equally distribute the effects of all possible sources of systematic and random bias and, as a result, to increase one's confidence that the obtained results are due to treatment. Control should not be lightly discarded or dismissed simply because an evaluation is conducted in a realistic situation. Control should be relinquished only hesitatingly, and as much control as possible should be exerted even if the evaluation takes place in the fleet or real world.

Types of Experiment

Campbell and Stanley (1966) differentiated four kinds of experiment: controlled, quasi, natural, and pseudo. Only two types, controlled and quasi, are of concern here. Both depend on the manipulation of variables.

In quasiexperimental research, the prefix suggests that the manipulation is left to nature. The manipulation occurs with minimum man-made design but allows for meaningful interpretations of the consequences, e.g., the effects of ash on crops can be assessed by sampling at various distances and directions from Mount St. Helena.

Designs

A great deal of literature is available on appropriate experimental designs. The interested reader is referred to Winer (1971), Kirk (1968), and Hays (1973) for treatments of experimental designs and applicable statistical techniques.

Controlled Experiments

Controlled experiments represent a "preferred" class of empirical research. In many ways, the preferred status is well deserved, although at times it has the appearance of an obsession that blinds some to its limits. Experiments are preferred for a number of reasons. When used properly, they add strength to what can be said with confidence about a situation under investigation and make it possible to rule out all or most of the obvious, alternative explanations.

This potential of controlled experiments is the primary reason for their status. The potential starts with a basic building block, referred to by some as a factor (Stanley, 1973), and by others as an independent variable (D'Amato, 1970), or a treatment (Nunnally, 1975). Each building block necessarily demands

at least two levels, values, cells, or conditions of the independent variable and represents at a minimum the presence or absence of the variable of interest. Combining and recombining basic building blocks allows for designs that range from the very basic to those of high complexity. The execution of increasingly complex designs is facilitated by the availability of good statistical models and the computer.

The controlled experiment's methods lead to analysis by a variety of linear statistical models. These allow the results to be decomposed into four uncorrelated, additive components: (1) treatment effects, (2) interactions, (3) interaction between subjects and observations in a repeated measures design, and (4) a residual component attributable to experimental error. Coupling the statistical models with the computer produces statistical machinery that some feel is overly elegant given the quality of the measurement data. Indeed, it is possible that the wide availability of computers and computer based statistical packages has proliferated the performance of complex and possibly ill-considered experiments--the constraint of the time and effort needed to analyze data being removed.

Identifying Features

Controlled experimental decision aid evaluation can be characterized as having a number of distinguishing features. The features includes: (1) direct control and manipulation of characteristics of the aid or the conditions of its use, (2) dependent variables that are indexed to the condition manipulated, (3) control of all relevant and confounding variables, and (4) random assignment of subjects to various conditions.

Items of Concern to Aid Evaluation

Within this context, a number of comments may be made about the design of experiments performed for decision aid evaluative purposes.

- Dependent variables should be linked directly to the independent variables. If the independent variable is various conditions of aid use, the dependent measure(s) should be some directly linked variable such as the number of hostile forces destroyed or amount of own force preserved.

- Subjects should be selected whose characteristics match or closely match those of the anticipated user. If the aid is designed for use by fleet aviation personnel, subjects with fleet air operational experience should be sought. Subject experience may also be treated as an independent variable.

- The problem solving scenario should resemble an opertional problem as closely as possible. This realism adds to the acceptability of the results to operational personnel.

- Problems at varying difficulty levels should be employed. Problem difficulty may be treated either as an independent variable or may be confounded.
- The use of some decision aids depends on familiarity with the operation of computer terminals and knowledge of the mechanics of employing the aid. Full training in these aspects is required if the results are not to be influenced by these variables.
- In order to avoid inadvertent influence of the evaluator on the results, the evaluator should be removed from the testing situation. Automatic recording of responses and automatic scenario presentation can do much towards eliminating subtle evaluator influences.

Quasiexperiments

Controlled experimental methods have been widely used to evaluate operational decision aids within the operational decision aid program of the Office of Naval Research. As previously indicated, the methods were extensively utilized within this program by both the aid developers and by independent evaluators (Siegel & Madden, 1980; Madden & Siegel, 1980). However, one problem with those studies is that the results were not readily generalizable to the fleet. The lack of realism or fidelity within the laboratory setting employed was the primary detriment to generalization of the findings. Quasiexperiments attempt to overcome this shortfall. They allow an opportunity to increase realism (and hence, generalizability of results) without serious jeopardization of necessary controls.

One feature which distinguishes "true" experiments from quasiexperiments is concerned with the type of independent variable manipulation involved. In true experiments, the independent variable is manipulated by the evaluator over one or more levels while other conditions are controlled or held constant. Quasiexperiments, on the other hand, do not rely on evaluator manipulation over levels and control of other variables which might affect results. Consider the following four evaluation schemes:

1. Group A: Normal Conditions → Introduce Aid → Calculate Difference

2. Group A: Normal Conditions → Introduce Aid → Calculate Difference
   → Remove Aid → Calculate Difference

3. Group A: Normal conditions
   Group B: Use Aid
   } → Calculate Difference

4. Group A: Use Aid → Remove Aid → Use Aid
   Group B: Normal Conditions → Normal Conditions
   → Normal Conditions
   } → Calculate Differences

While there is a "control" group in several of these schemes, there is no deliberate attempt to hold conditions constant or to equate the various groups. Moreover, subjects may enter or leave a group during the course of the evaluation.

Finally, quasiexperiments often rely on correlational methods and measures of association. On the other hand, "true" experiments rely on difference measures.

Dependent Variables

One curious observation about quasiexperiments is that they are associated with multiple criteria. In controlled experiments, one (or at the most two) dependent measures are typically employed. In quasiexperiments, more dependent variables are often available and employed "in order to weave a net of circumstantial evidence regarding the 'reality' of observed findings" (Nunnally, 1975).

In experimental evaluative research, the dependent variables are usually some index of decision quality. The dependent measure of decision quality is usually marked as the end point and terminates the scenario. This arrangement facilitates the data collection but might not be coincident with the arrangement found in the fleet. Rather, in the fleet, an aid might be used iteratively to work out courses-of-action in a volatile, rapidly shifting situation. In such a situation, a number of criteria might be needed to determine how and to what extent an aid facilitates the operational requirements.

A further reason for multiple criteria in quasiexperiments relates to the possibility of lack of control over relevant confounding variables. To the degree that control is not possible, more evidence is needed to be able to make convincing arguments. For example, if evidence from various sources can be demonstrated to covary, then the convergence can be used to argue for or against the usefulness of the aid.

Partial Correlation

In a quasiexperiment, random assignment of subjects to groups and treatments may be the exception rather than the rule. Subjects are often assigned to groups by virtue of circumstances. Partial correlation is appropriate when groups possess characteristics that may influence or be related to the variables of the decision process.

Uncovering Spurious Relationships

Properly used, partial correlation becomes an excellent technique for uncovering spurious relationships. A spurious correlation is a relationship between

two variables (A and B) in which the relationship is the result of the fact that A and B vary with some other variable, C. Should the effects of C be controlled or held constant, then A and B may no longer covary. Accordingly, a partial correlation between two variables is one that nullifies the effects of a third variable on both of the remaining variables being correlated.

Example of a Spurious Relationship

The correlation between decision performance and aptitude of personnel where age is permitted to vary would be higher than the correlation between decision performance and aptitude in a group at constant age. The reason is obvious. Older personnel have more operational experience. Age is a factor that enhances the strength of the correlation between decision performance and aptitude. If an evaluator wishes to know the correlation between decision performance and aptitude with the influence of age ruled out, the evaluator can control for age in his selection of subjects. However, the partial correlation technique enables the evaluator to accomplish the same result without forming two equivalent age groups.

Intervening Variables

Another possible application of partial correlation exists because of its ability to aid the evaluator in his search for intervening variables. While there is no mathematical difference between the computation of partial correlations designed to locate spurious relationships and those used to identify intervening variables, the conceptual issues are different. The search for intervening variables is highly related to the issue of causality. For example, an evaluator may wish to say that variable A leads to variable B, which in turn leads to variable C.

Consider a hypothetical study in which an evaluator is concerned with the contribution of personnel characteristics and decision aid effectiveness to mission success. Given the matrix of correlation coefficients shown in Table 3, the evaluator might hypothesize that: (1) the contribution of personnel characteristics and aid effectiveness are direct, or (2) the major portion of the correlation is due to the indirect relationship between aid effectiveness and personnel characteristics. A solution to this problem can be reached by computing: (1) the partial correlation between aid effectiveness and mission success, (2) the partial correlation between personnel characteristics and mission success, and (3) an examination of the zero order correlation coefficients. The conclusion reached in the case of the Table 3 data is that aid effectiveness transfers to personnel characteristics which then transfers to mission success.

Table 3

Hypothetical Matrix of Correlations

	MS	AE	PC
Mission Success (MS)	1.00	.50	.42
Aid Effectiveness (AE)		1.00	.68
Personnel Characteristics (PC)			1.00

Locating Relationships

Another problem is locating relationships where none appear to exist. Here, again, the mathematical methods are the same, but the conceptual issues are different. Suppressor variables can act to hide or suppress true relationships. Variable A may show no relationship to variable B because variable A is negatively related to variable C, which in turn is positively related to variable B. Accordingly, variable A may be expected to be positively related to variable B when one controls for the effects of variable C. One may employ partials to provide insight into such questions (Nie et al., 1975).

Statistical Issues in Partial Correlation

When only one variable is held constant, it is customary to speak of a first order partial correlation. When two variables are held constant at the same time, the coefficient is called a second order partial correlation. Actually, there is no limit on the number of variables that can be held constant. However, in actual practice, fourth order and higher partials are seldom seen.

In one sense, the use of partial correlation is a statistical substitute for experimental control. Linearity of relationships is assumed. When the relationship is nonlinear, only the linear component of the relationship is partialed out.

Multiple Regression

Multiple regression enables the evaluator to examine systematically the relationship between a dependent (criterion) variable and a set of independent (predictor) variables. The specific question of interest to the evaluator may be: What combination of aid characteristics best predicts the criterion?

In a standard multiple regression, all specified predictor (aid characteristics) variables are entered into the regression equation in a single step. Alternatively, each of the predictors can be entered into the equation one by one in either an order specified by the evaluator or on the basis of statistical criteria.

A multiple regression prediction equation takes the four of $Y: = a + bX$, $+ b_2X_2 + \ldots + b_nX_n$, where Y is the estimated or predicted value of aid effectiveness, X_1, $X_2, \ldots$, X_n are the aid characteristics, b_1, $b_2, \ldots$, b_n are weights by which the values of the aid characteristics are multiplied, and a is the constant added.

Example of Application

An example of the application of regression analysis in a design in which one independent variable is continuous and another is categorical is the situation in which an evaluator is interested in the effects on decision accuracy of: (1) a decision aid, and (2) the preparation level of the decision makers. Another question of interest to the evaluator might be whether or not there is an interaction between the two groups, one using the decision aid and the other not using the aid. Within these groups, subjects are allotted 50, 100, 150, or 200 minutes of training time. At the end of a training period, a mission simulation is conducted to collect data. Note that while aid vs. no aid is a categorical variable, preparation time is continuous. Such a study will produce two regression lines. Two questions may be asked about the regression lines. First, are the regression lines parallel? Equality of slopes means that the effect of preparation time is the same whether or not the decision aid is employed. Second, is the elevation of the regression lines equal? Equality of slopes and intercepts means that a single regression line fits the data for both groups. In such a case, it may be concluded that the use of the aid has no favorable effect. If, on the other hand, the regression coefficients (b's) are equal while the intercepts (s's) are not, the results indicate that the aid affects performance along the entire training time continuum.

Cross-Lagged Correlation

In a decision aid evaluative study, the evaluator may wish to know whether or not any decision improvement evidenced in an aided condition is actually caused by the use of the aid. Correlational measures show association but not necessarily causation. However, causal relationships may be established on the basis of correlational studies. This approach has been classified as quasiexperimental by Campbell and Stanley (1966).

Logic of Cross-Lagged Correlation

The following discussion of cross-lagged correlation is based on the discussion of Cook and Campbell (1979). In Figure 3, A and B represent two variables. Each variable is longitudinally measured at two points in time. The correlations, rA_1B_1 and rA_1B_2, and the retest correlations rA_1A_2 and rB_1B_2, provide a framework for interpreting the cross-lagged correlations, rA_1B_2 and rB_1A_2. If A is a cause of B, it can be expected that rA_1B_2 will be greater than rB_1A_2. In the figure, there is a clear $A \rightarrow B$ causation. Increases in A_1 cause increases in B_2. Equivocality of interpretation is avoided because rA_1B_2 is also greater than rA_1B_1 or A_2B_2 and because the latter two are approximately equal.

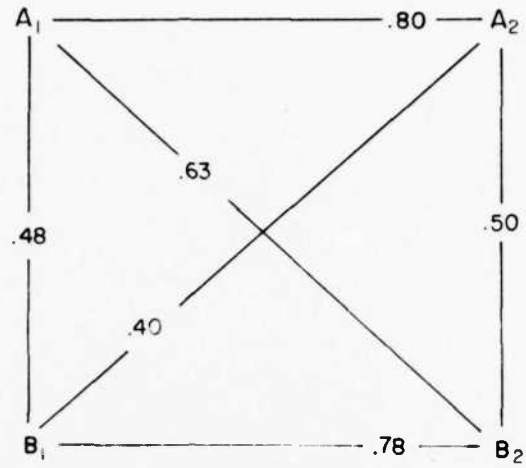


Figure 3. Cross-lagged correlation framework.

Example

Figure 4 presents hypothetical cross-lagged correlational data for ratings of satisfaction with an aid and use of the aid during operations. Two data collection points separated by six months in time are represented. There is a strong rA_1B_2 relationship, and rA_1B_2 (.70) is greater than rA_1B_1 (.42) and rA_2B_2 (.48). The results can be interpreted as indicating that satisfaction with an aid causes it to be used during operations.

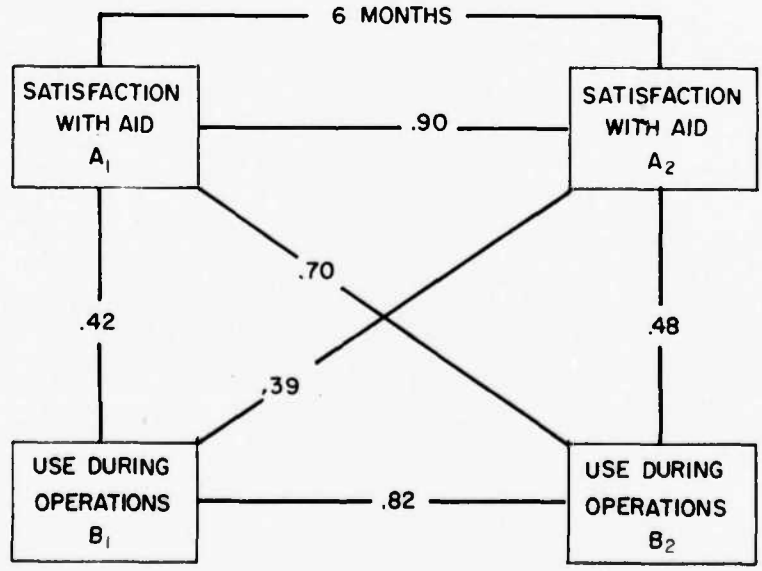


Figure 4. Hypothetical cross-lagged correlatation coefficients for satisfaction with aid and use of aid during operations.

Problems and Limitations

Kenny (1979) views cross-lagged correlation as a valuable tool for ruling out rival hypotheses of spuriousness. Kenny views the technique as largely an exploratory strategy for data analysis designed to uncover simple causal relationships. An experiment in which the causal variable is manipulated should follow a cross-lagged correlational finding.

Structural Equation Models

Structural equation models represent another way for answering the basic question asked by cross-lagged correlation--whether or not a given action was responsible for an observed event.

There are two types of structural equation models: recursive and nonrecursive. A recursive model is one in which all the causal linkages run "one way"; that is, no reciprocal relationship exists between two or more variables. The nonrecursive model is one in which two or more variables operate, such that "each affects and depends on the other" (Duncan, 1975) and act somewhat like a feedback loop.

Recursive models are the most widely used because their nature makes them easy to solve by standard least squares analysis. Nonrecursive models, on the other hand, unless specified very exactly, can be difficult to solve.

A recent conceptualization of how to solve structural equation models has resulted in the method of maximum likelihood estimations. Also, computer analytic packages such as LISREL, enable rapid, efficient estimation of the coefficients based on maximum likelihood. The coupling of the different approach to solving structural equations with a readily available computer package should lead to greater use of structural equation models. This is especially true because the maximum likelihood estimates allow for solutions to many structural equation models which in the past would have not been solvable.

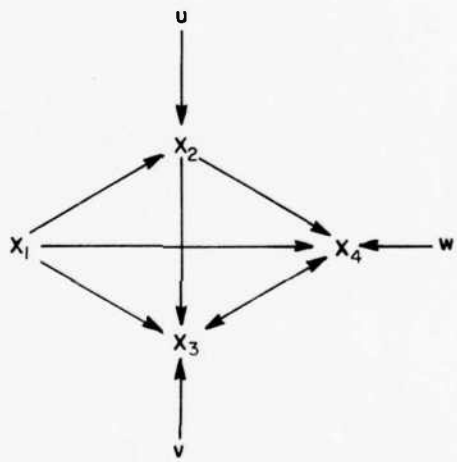
Developing Structural Models

Developing a structural model requires some understanding of the situation under study. It demands thinking about the problem in causal terms. One way to develop a structural model is to think of all the possible explanatory "causes," reject as many as possible, and then model the remainder. Assume that there are four dependent variables and a model is to be built that expresses the relationships among them. Also, assume that the variables can be arranged in an unambiguous causal ordering--X_1, X_2, X_3, and X_4. The ordering indicates that X_1 is not caused by any of the other variables; X_2 is not caused by X_3 or X_4, etc. Because X_1 has no cause among the variables, it is called exogeneous, while the remainder are called endogenous. A matrix of the possible effects of each variable can be developed, as below:

	Causal			
Effect	X_1	X_2	X_3	X_4
X_1	-	0	0	0
X_2	✓	-	0	0
X_3	✓	✓	-	0
X_4	✓	✓	✓	-

The 0 entries indicate that no causal relationship exists in a row-column intersect. As entered, the model is recursive.

These relationships can be illustrated as:



The structural equations for the model are:

$$X_1 = \text{exogeneous variable}$$

$$X_2 = b_{21}X_1 + u$$

$$X_3 = b_{32}X_2 + b_{31}X_1 + v$$

$$X_4 = b_{43}X_3 + b_{42}X_2 + b_{41}X_1 + w$$

where: b is a structural coefficient and indicates the influence that the "independent variable" has on the dependent variables, and

u, v, and w are unidentified disturbance sources and represent the variance in the dependent variable that cannot be accounted for by the structural coefficients.

The matrix, equations, and diagram all represent the same model. The merit of the model is measured by the variance that it can account for in a set of observations. Each equation is solved by using variances and covariances estimated from sample moments to estimate, in turn, the structural coefficients. The coefficients can then be used to determine R^2, the amount of variance accounted for by the model.

Using the squared multiple correlation to assess a model's power is one, but probably not the best way to evaluate the result. Insofar as structural equation models have other properties that make them valuable, the size of R^2 should not be a sole determining factor in assessing the model, e.g., the model might have heuristic value because it allows for a clear illustration of some relationships, or it might have value because it allows for the generation of important hypotheses.

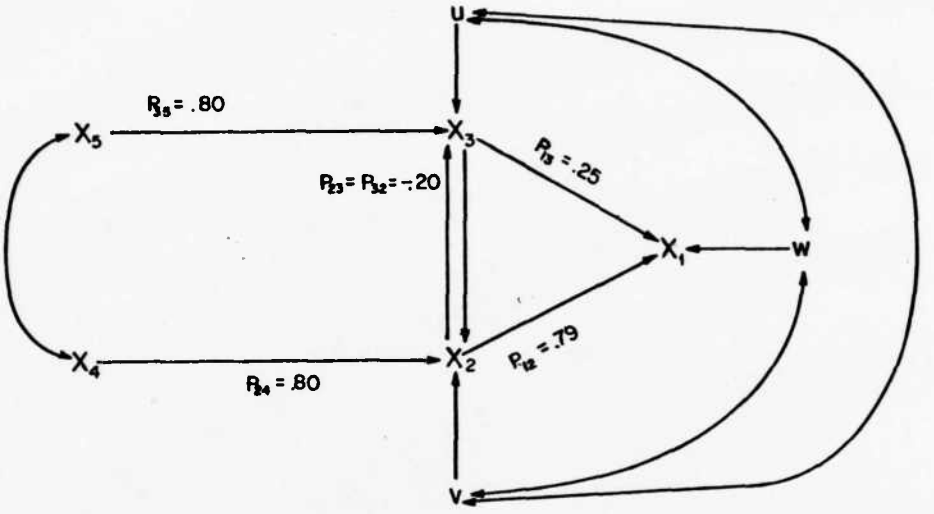
Application to Decision Aid Evaluation

At least two methods for employing the structural equation approach to decision aid development are evident. One is concerned with modeling the decision process itself. The other rotates around using data derived from the application of the analytical techniques (Chapter III) to determine what causes a given aid to possess merit.

Decision Process

The numerous variables that might have an effect in determining the outcome of the decision process provide a clear basis for developing a structural equation model. A structural equation model (Madden & Siegel, 1980) of the decision process for emission control in Navy task forces serves as an example. Five variables were selected for inclusion in the model. Three were treated as endogenous variables and were outputs of the decision aid: (1) tradeoff score, (2) surveillance score--indicative of degree of protection of the task force, and (3) information denial score--indicative of the targetability of the task force. The tradeoff score indicated the degree to which the other two scores are balanced. The final two variables were treated as exogenous--one symbolized the need for protection and the other a security need.

The structural model is illustrated as a diagram in Figure 5. Straight lines in the model indicate a causal relationship. The direction of the causality is indicated by the direction of the arrows on the straight lines. The bent lines containing bidirectional arrows indicate relationships between exogenous variables and between disturbance terms.



ENDOGENOUS VARIABLES
- X_1 = TRADE SCORE
- X_2 = SURVEILLANCE SCORE
- X_3 = INFORMATION DENIAL SCORE

EXOGENOUS VARIABLES
- X_4 = NEED FOR PROTECTION
- X_5 = NEED FOR SECURITY

DISTURBANCES
- u
- v
- w

STRUCTURAL EQUATIONS

$$X_3 = P_{35} X_5 + P_{32} X_2 + u'$$

$$X_2 = P_{34} X_4 + P_{23} X_3 + v'$$

$$X_1 = P_{13} X_3 + P_{12} X_2 + w'$$

Figure 5. The tradeoff structural equation model for one operational decision aid.

The exogeneous variables are not the concern of the model and are conceptually distant, but necessary, components. The first exogeneous variable, X_4, represents the need to protect the task force, and it was assumed that the effect of this need was to maximize surveillance coverage. The second exogenous variable, X_5, represents a need for security, and it was assumed that this need resulted in a maximization of information denial. Both exogeneous variables are conceived as latent variables which cannot be measured directly and must be estimated. The premise behind these variables is that they represent two divergent responses to a state of primary motivation. This primary motivational state, which is not incorporated within the model, is assumed to be a function of a number of factors, such as the mission of the task force, the psychological makeup of the task force commander, relevant intelligence, and so forth.

Each exogeneous variable has a direct effect on one endogeneous variable. The "protection need," X_4, has a direct causal effect on X_2, the measure of surveillance coverage (the surveillance score), forcing it toward a maximum value. The "security need," X_5, has a direct causal effect on X_3, the measure of information denial, forcing it toward its maximum.

Obviously, maximum values of the X_2 and X_3 variables cannot lead directly to the balance the tradeoff requires. Therefore, some other causal linkage must be assumed which will modify these maximizing tendencies. The causal linkage between X_2 and X_3 represents the point in the model where reciprocating activities result in a modification of each. The paths of each, the effect of X_2 and X_3 and that of X_3 and X_2, are represented in Figure 5 by the double lines between them. Each arrow represents an effect in the opposite direction. It is further assumed that X_2 and X_3 both affect X_1, the tradeoff score (the measure of achieved balance).

The model attempts to account for the variance in each of the endogenous variables, X_1, X_2, and X_3, as functions of causal relationships. However, total variance can never be completely accounted for by a model. The relationships are therefore assumed to be attentuated by disturbance terms. The disturbances shown in Figure 5 (u, v, and w) represent other correlated sources of variance in the endogenous variables not specified in the model (Duncan, 1975).

The model, which is nonrecursive and fully identified, may be represented by three structural equations:

$$X_3 = p_{35}X_5 + p_{32}X_2 + u'$$

$$X_2 = p_{24}X_4 + p_{23}X_3 + v'$$

$$X_1 = p_{13}X_3 + p_{12}X_2 + w'$$

where the p's are the structural or path coefficients and where because of the correlation in the disturbance terms:

$$u' = p_{3u}u + p_{32}p_{2v}v$$

$$v' = p_{2v}v + p_{23}p_{3u}u$$

$$w' = p_{1w}w + p_{13}u' + p_{12}v'$$

In structural equation models, a simplifying assumption is that the path coefficients for the direct effect of the disturbances equals one (Duncan, 1975). Therefore, the disturbances terms can be rewritten:

$$u'' = u + p_{2v}v$$

$$v' = pv + p_{3u}u$$

$$w' = w + p_{13}u' + p_{12}v'.$$

Implications

The model suggests that the composition of the aid is such as to support the motivational and the cognitive needs for developing an effective plan. In addition, byproducts of the model development can be meaningful for an evaluator, e.g., the structuring of the model and the testing of various alternatives permits one to identify important sources of information.

The final implication of ascertaining the causal relationships is that the data can be used as an additional argument about usefulness. This evidence may very well be unique. Subtle differences in the weight that a user places on a variable can be brought out by the appropriate application of the structural equation model approach.

Use of Data from Analytic Techniques

The second area of application of structural equation models is concerned with analyzing the relationships between the dependent variables representing the output of the analytic methods and the decision quality produced by an aid.

The variables used in the subsequent exposition are derived from the Shalit Perceptual Organization and Reduction Questionnaire and the Polydiagnostic Method. From the Shalit technique, the reciprocal of the reduction index (Red I), the emotionality index (E), and the intensity index (I) are used, while three hypothetical variables (V_1, V_2, V_3) are used that can be thought of as expressing qualities found to describe an aid's "personality" as indicated by the Polydiagnostic Method.

Two models will be developed. Both are designed to assess "decision quality."

Multiple Indicator Model

A multiple indicator model is one in which one or more variables are indexed by any number of indicator variables. The indicator variables are represented as functionally related to the indexed variable(s). How the models can be differently developed and the subsequent differences in their illumination of reality is demonstrated by two multiple indicator models--Model I and Model II. Both models are presented in Figure 6.

Both variations assume that a latent variable, L, conceived as a primary motivational state directly affects qualities V_1, V_2, and V_3 measured by the Polydiagnostic Method. The qualities, in turn, have no effect on other variables. They are represented as multiple indicators of the current state of the latent variable.

In both model variations, the latent variable affects the E and I scores derived from the Shalit technique. E and I then directly affect the output from the aid, A, B, and C.

Beyond this point, the models differ. Model I indicates aid related variables as affecting two dependent variables in independent and parallel processes. The variables affected are the reduction index and decision quality. Accordingly, the model conceives of both the perceived structural organization and the final decisions as being multiple indicators of the intensity, emotionality, and the outputs from the aid.

Model II, on the other hand, suggests that the aid's output directly affects the user's perceived structural organization, the reduction index, and this, in turn, functions as the immediate cause of the decision quality. The process here then is dependent and serial.

"Reality"

The two models present different pictures of "reality." Model I suggests that the perceived structuring of the environment is a function of the aid use but has no bearing on the decision quality. However, Model II indicates that decision quality is a function of perceived environmental structure. The differences between the models demonstrate that rather different interpretations of events could be incorporated in different structural equation models and, accordingly, provide a reasonable basis for testing why an aid functions as it does.

Limitations

No model may be thought to be an expression of a fundamental theory. Models are meant to be descriptive. They may be indicative of a plausable causal ordering of variables. Much is left out of any model. In addition, causal models are often simplified by making them recursive. Such oversimplification may represent a highly unlikely state of affairs, given the interactive nature of the operational environment in which a decision aid is used.

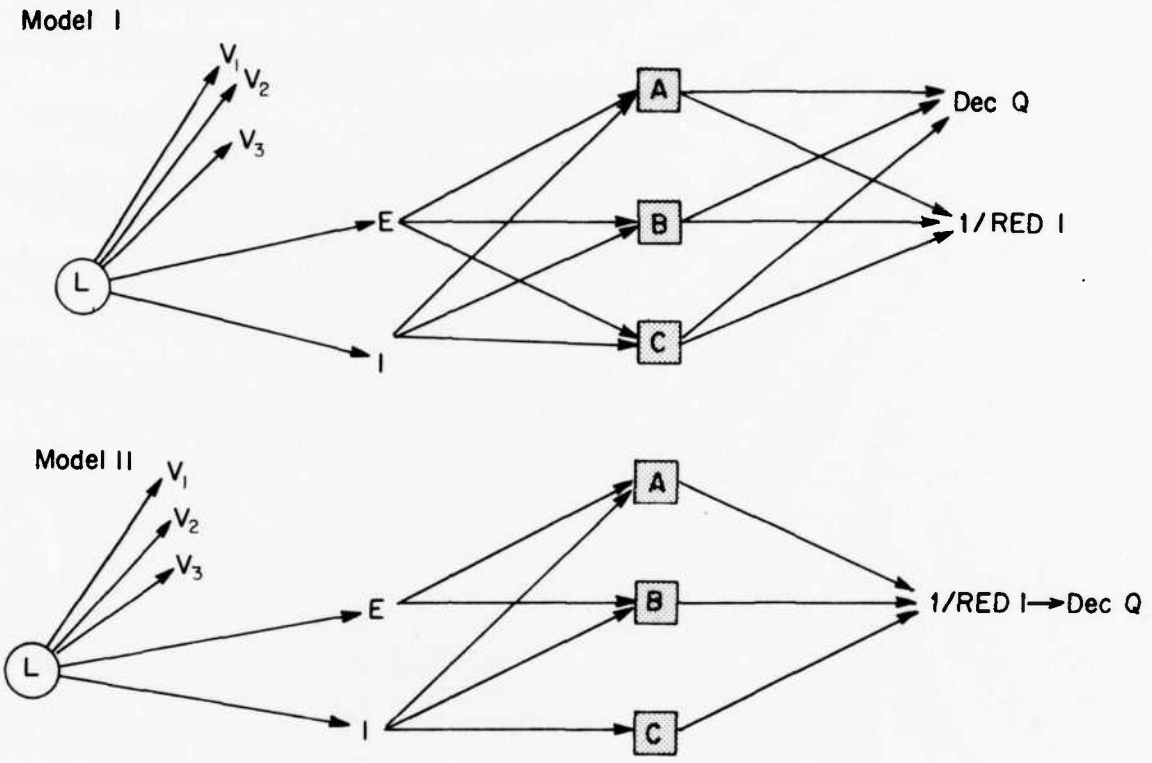


Figure 6. Two possible multiple indicator models using the output of two analytic techniques.

1/RED I	=	reciprocal of reduction index
E	=	emotionality index
I	=	intensity index
V_1	=	quality 1, hypothesized
V_2	=	quality 2, hypothesized
V_3	=	quality 3, hypothesized
Dec Q	=	Index positive related to decision quality
A	=	display found by matrix analysis to be highly related to user's decisions
B	=	display found by matrix analysis to be highly related to user's decisions
C	=	display found by matrix analysis to be highly related to user's decisions
L	=	latent variable, exogenous

Time Series Analysis

The time series design relies on repeated measurement over time of the behavior or process under investigation. The method has traditionally be used when some event occurs which may affect the situation under study. The investigator looks for an association between changes in trend of the situation under study and the occurrence of an event, i.e., introduction of a decision aid. The time series design seems particularly applicable to aid evaluations conducted in the fleet.

Simple Interrupted Time Series

The most basic time series evaluation design requires one experimental group and multiple observations before and after treatment (introduction of a decision aid). The design is diagrammed below. Here, X represents the time of the treatment introduction.

O_1	O_2	O_3	O_4	X	O_5	O_6	O_7	O_8

Figure 7 presents, as an example, ficticious data acquired before and after the introduction of a decision aid. A major advantage of this time series design is that it allows assessment of the maturational trend prior to, as well after, the intervention.

Interrupted Time Series With No Treatment Control Group

If one considers the addition of a no treatment control group to the prior design, the resulting quasiexperimental evaluation design is diagrammed below:

Group A:	O_1	O_2	O_3	O_4	X	O_5	O_6	O_7	O_8
Group B:	O_1	O_2	O_3	O_4		O_5	O_6	O_7	O_8

Hypothetical results from such a study are shown in Figure 8. Figure 8 suggests that the introduction of the aid caused improvement in decision accuracy because the trend lines diverge after introduction of the aid.

Interrupted Time Series With Removed Treatment

The interrupted time series design with removed treatment essentially involves two interrupted time series. The design is diagrammed below:

O_1	O_2	O_3	O_4	X	O_5	O_6	O_7	O_8	X	O_9	O_{10}	O_{11}	O_{12}

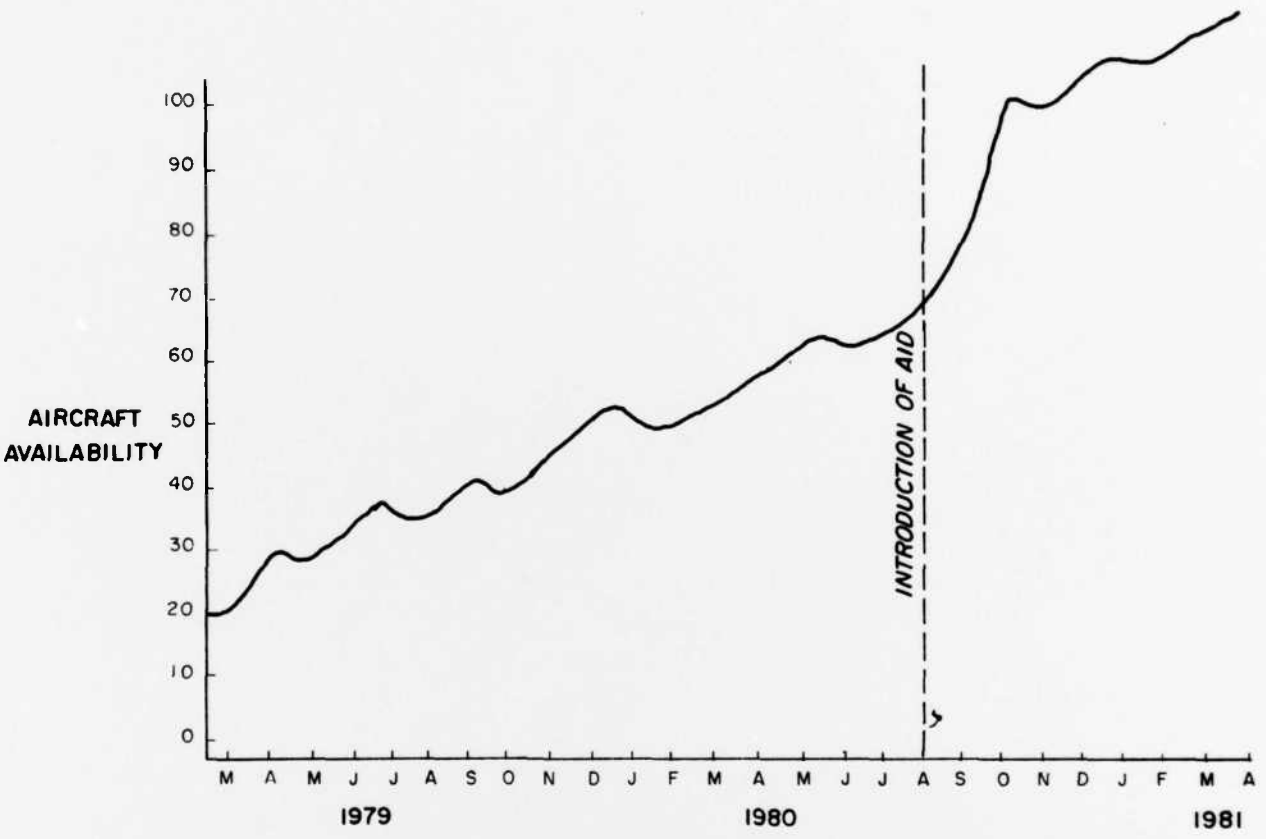


Figure 7. Change in aircraft availability as a result of introducing a maintenance decision aid (hypothetical data).

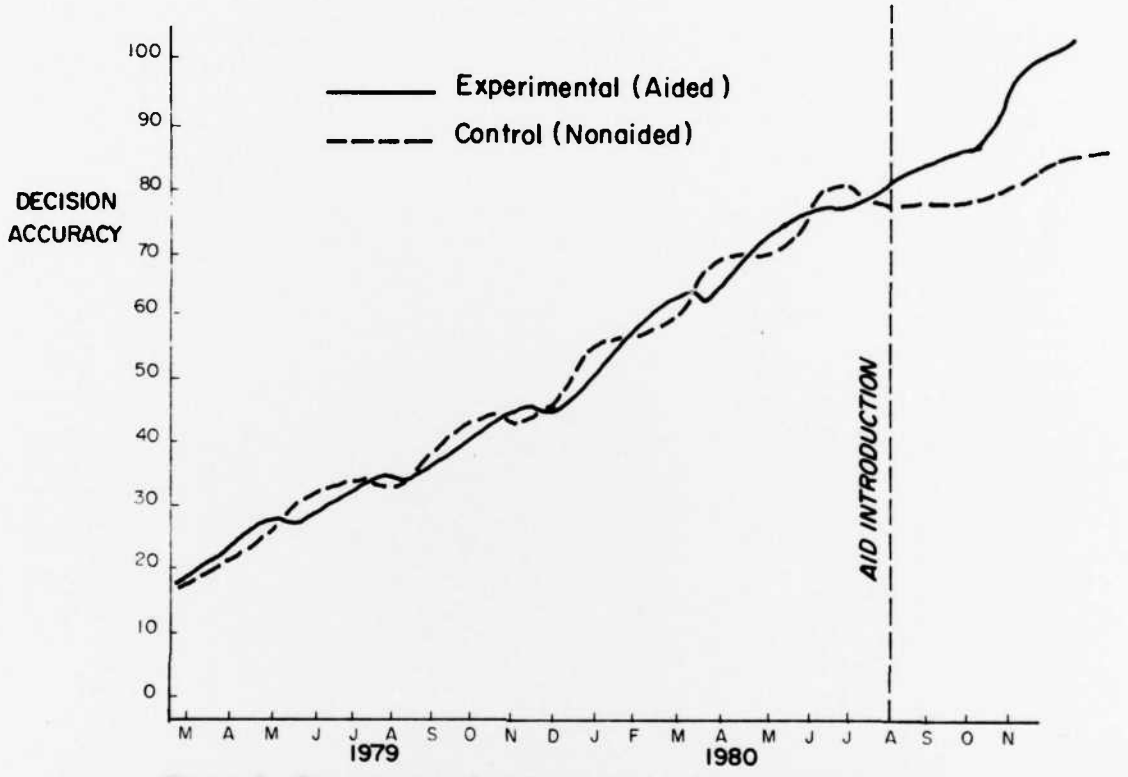


Figure 8. Divergence in decision accuracy for experimental and control groups after introducing the decision aid to the experimental group (hypothetical data).

The first stage, from O_5 to O_8 is designed to permit assessing the effects of the presence of the treatment (introduction of the aid) and the second, from O_9 to O_{12}, is designed to assess the effects of removing the treatment (removal of the aid). One might hypothesize that the slope would change in one direction between O_5 and O_8 and then change in the opposite direction between O_9 and O_{12}. Since there are two separate time series, the overall design has somewhat more strength than a simple time series.

Interrupted Time Series With Switching Replications

Given two samples, each of which receives the aid at different times, one group can serve as a control for the other. One group serves as a control and later receives the aid; then, the original treatment group serves as the control. The design is diagrammed below:

Group A:	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	X	O_9	O_{10}
Group B:	O_1	O_2	O_3	X	O_4	O_5	O_6	O_7	O_8	O_9	O_{10}

The logic of this design is that external validity is enhanced when an effect can be demonstrated with two samples in two settings at different periods of time.

Statistical Analysis of Time Series Data

While several techniques may be employed in the analysis of change over time, regression analysis is the one most commonly employed. Two fundamental types of data are made available by the time series design: (1) cross sectional data including observations on data at a given point in time, and (2) time series data, in which one has a set of observations over a series of time points. One can employ techniques based on regression analysis for either type of data. However, one critical difference between cross sectional analysis and time series analysis is that, for the latter, it is critical that the data be processed in order of the time periods involved.

Auto regressive integrated moving average (ARIMA) analysis developed by Box and Jenkins (1976) represents the most widely used time series analytic method. Time series models typically have two components. The first component describes the systematic behavior of a time series and is called the deterministic component. The second component is called the stochastic component because it describes an underlying process of unobserved errors that make the observed time series less predictable. A major goal of the analysis is to discover the structure of the systematic part of the stochastic component and represent it as an equation.

Limitations of Time Series Designs

Time series designs are usually based on interventions at a fixed point in time. Unfortunately, an intervention at a fixed point in time may diffuse slowly through the population. Treating a diffusion curve as though it is a step function can create problems of interpretation. When it is anticipated that an intervention will disseminate slowly, it is desirable to collect data designed for the purpose of describing the diffusion process (Box & Jenkins, 1976).

Textbooks dealing with the statistical analysis of time series designs suggest different rules of thumb for the number of time points required. About 50 observations are generally considered as sufficient for estimating the structure of the correlated error.

CHAPTER V

A SYSTEMS APPROACH TO DECISION AID DEVELOPMENT AND EVALUATION

The purpose of Chapter V is to outline a procedure for applying general system and system engineering approaches to the evaluation of decision aids. This requires that decision aids be conceived as man/machine systems or subsystems. Taking the systems perspective rests on the view that decision aids and their users are components of subsystems that must fit into the larger collection of subsystems that compose a total system. The major subsystems of the larger system could be classified by function such as command and control, communication, information, radar, sonar, etc.

Developmental and Evaluational Scheme

The hallmark of a good system is that it does what it was intended to do. Arriving at this state-of-affairs is not a chance happening. Rather, it is achieved by careful design and test of the system, its features and subfeatures. This is represented by a cycle of analysis, design, development, implementation, test, and redesign. Each stage of the cycle is completed when some prespecified criteria are demonstrated to be met. Then, the next level is entered until finally, after having proven its capabilities at each development stage, the system does what is was intended to do and full operational deployment is implemented.

Note that within the cycle, the output of each stage represents the input to the next stage.

Following the principle of a continuous process with the output from each stage representing the input to the next and full test at each stage of the development cycle, the developmental cycle for a decision aid may be conceived as shown in Figure 9. The figure is read from the left to the right, with the potential evaluative techniques which are applicable at the conclusion of each stage shown in the shaded box following each stage. The five stages in the development of decision aiding subsystems are identified as: (1) system conception, (2) system definition, (3) system design and development, (4) validation, and (5) operational evaluation.

Stage 1--System Conception

The decision aid design, development, and implementation cycle starts, in this context, with a conceptual phase. The output of this phase is a system concept. To derive the decision aid system concept, a variety of sources may be consulted and a variety of analytic techniques employed. Possibly, the most important of these is a system/mission analysis in terms of the goals of the

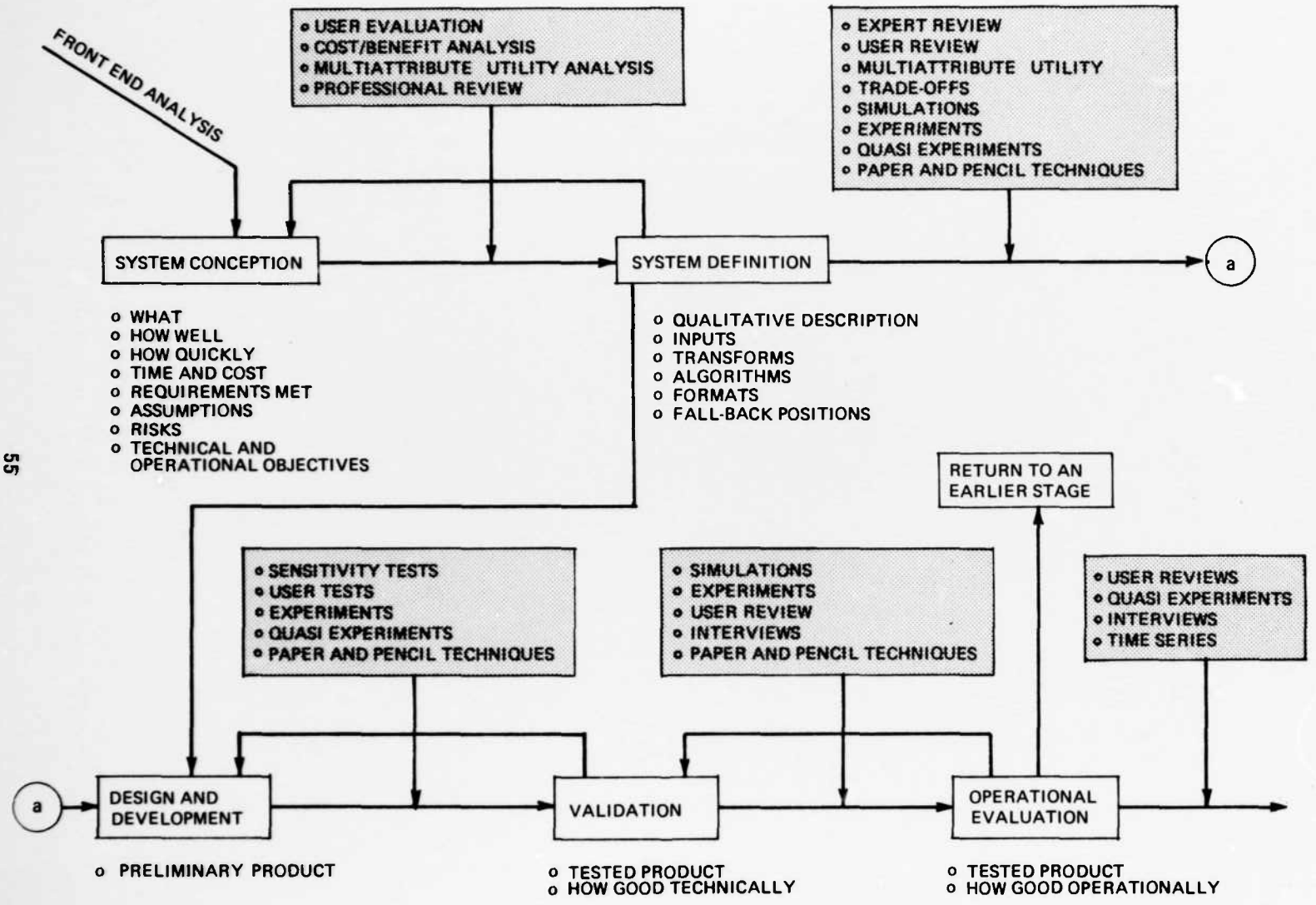


Figure 9. Aid developmental sequence.

mission to be supported by the aid, the current decision aiding needs and their importance, the current operator functions (unaided), the operator functions which require aiding, the required decision accuracy, and some quantification of current sources of errors and how the errors might be reduced by aiding. The result of the work of this stage is a full concept which is sufficiently detailed to allow one to know with some degree of exactness what the aid will do, how quickly it will do it, what the developmental time and cost will be, what requirements the aid will meet and what requirements it will not meet, the assumptions and risk areas, and, most importantly, the aid's technical and its operational objectives and how the aid will accomplish these objectives.

The concept can then be subjected to evaluation through a number of evaluative techniques. These include, but are not limited to: interviews with user personnel, cost/benefit analysis, multiattribute utility analysis, and review by other professionals. Minimal considerations during such reviews are: internal consistency, generality, cost of development and use, content validity, richness of output, construct validity, and completeness.

In the case of conflicting or alternative concepts, a tradeoff analysis can be completed to allow the choice of a preferred concept.

Stage 2--System Definition

The system definition stage commences once the concept has been refined on the basis of the prior evaluation(s) and fully agreed on. During the definition phase, the aid's concepts are hardened, the aid is fully specified in terms of the input data it will accept, and the output data it will provide. The format of each output display is specified, and the internal structure is developed. This includes all necessary computer logic, algorithms, and representations. The programming language is also selected during this stage on the basis of compatibility with the anticipated host computer system.

The end result of the work of this phase is a concrete description of the aid, its input data and how the input will be converted to outputs, what the input and the output formats will be, and fall-back positions for any problem areas.

The evaluative techniques available for use at the conclusion of this stage are: expert review, user review, multiattribute utility analysis, cost analysis, and tradeoff analysis. Mock-up and/or simulation studies can serve to refine the user interface. These can employ the experimental and the quasiexperimental methods described in earlier sections of this report. Also, several of the paper-and-pencil analytic methods, described earlier, can be employed for evaluative purposes at this developmental stage.

Internal consistency, generality, cost of use, content validity, richness of output, and merit ordering must again be considered at this stage.

Stage 3--System Design and Development

The output from the definition stage represents the input to the design stage. In the design stage, the aid is programmed in accordance with the system definition, and a preliminary product results.

This preliminary product is subjected to a number of tests. These are largely related to the sensitivity of the aid. In these tests, input parameters are varied, and the effects of this input variation on output are examined. Are the effects realistic? Do they seem to be in the proper direction and of the proper magnitude? Are any illogical results obtained? Are the results continuous? Particular attention is paid at this juncture to indifference to trivial aggregation, correct directionality over the entire range, internal consistency, and satisfaction of specifications. Limited user tests are also suggested at this stage. These tests help to uncover problems associated with the user-aid interface. If operating instructions are to be supplied with the system, these should also form a part of the user tests. And, as before, the paper and pencil evaluative techniques and interviews can be employed to allow an additional verification of the operator-aid interface.

Stage 4--Validation

The validation stage starts with the decision aid as revised at the conclusion of the development stage and ends with a set of statements about the extent to which the final aid has achieved the purposes established during the system conception stage. Regardless of context and specifics, the purpose of a decision aid is to help a decision maker to make decisions. Accordingly, validation implies a quantitative measure of the extent to which this goal is achieved. Other questions involved in validation may involve why the aid works, what attributes of the aid make it work, and how can it be made to work better?

To these ends, some type of at least moderate fidelity simulation of the context in which the aid will be used as well as high fidelity simulation of the aid itself are basic. The experimental and the quasiexperimental methods, described earlier, are appropriate here, but a number of the other techniques are appropriate for collecting supplementary data relative to other questions of interest.

Validation is not necessarily a unitary event. Several validational studies may be necessary to acquire all important information. And, even if the results of an initial validation are positive (or negative), cross validation is usually required.

The validation results will state, to a greater or a lesser extent, how well the decision aid meets its technical objectives. If the technical objectives are met, then an operational evaluation is in order.

Stage 5--Operational Evaluation

During the operational evaluation stage, the aid is installed on an operating system and tried during Navy operations. Here, the quasiexperimental methods, user evaluation, and the interview methods, described earlier, will be most appropriate. The difference between the objectives of the validation stage and the operational evaluation stage is that the validation stage is primarily concerned with whether or not the aid meets technical objectives, while the operational evaluation is concerned with whether or not the aid serves operational objectives.

After operational evaluation, the aid may enter limited, and then extended, use or some refinements may be indicated. If refinements are indicated, the required redesign is implemented. This may require a return to one of the earlier stages. After redesign, operational evaluation is performed again.

REFERENCES

Allen, P.S., Bennett, E.M., Kemler, D.K., & Carter, W.K. Forced-choice ranking as a method for evaluating psycho-physiological feelings. Dayton, Ohio: Wright Air Development Center, 1959.

Barclay, S., Paterson, C.R., Randall, L.S., & Donnell, M.L. Decision analysis as an element in an operational decision aiding system, Phase V. McLean, Virginia: April 1979.

Bennett, E.M. The polydiagnostic method. Journal of Psychology, 1956, 42, 207-215.

Bennett, E.M., Kemler, D.K., & Allen, C.S. The polydiagnostic method of multiple forced-choice rankings in design analysis. Wright-Patterson Air Force Base, Ohio: Aero Medical Laboratory, December 1958.

Box, G.E.P., & Jenkins, G.M. Time series analysis. San Francisco-Holden-Day, 1970.

Campbell, D.T. Factor relevant to the validity of experiments in social sittings. Psychological Bulletin, 1957, 54, 297-312.

Campbell, D.T., & Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand-McNally, 1966.

Cook, T.D., & Campbell, D.T. The design and conduct of quasi-experiments and the experiments in field settings. In M.D. Dunnette (Ed.), Handbook of Industrial and Organizational Psychology. Chicago: Rand McNally, 1976.

D'Amato, M.R. Experimental psychology: Methodology, psychophysics, and learning. New York: McGraw-Hill, 1970.

Duncan, O.D. Introduction to structural equation models. New York: Academic Press, 1975.

Edwards, W. Social utility. Paper presented at the Symposium on Design and Risk Analysis. Annapolis, Md.: American Society for Engineering Education and American Institute of Industrial Engineers, 1971.

Feigl, H. Principles and problems of theory construction in psychology. In W. Dennis (Ed.), Current trends in psychological theory. Pittsburgh, Pa.: University of Pittsburgh Press, 1951.

Fischl, M.A., & Siegel, A.I. Further research into and validation of the analytic profile systems for visual display evaluation. Wayne, Pa.: Applied Psychological Services, 1970.

Glenn, F. Informal evolution of the ASTDA engagement model. Willow Grove, Pa.: Analytics, November 1978.

Glenn, F., & Zachary, W. ASTDA user's guide. ONR Report 1344-A. Willow Grove, Pa.: Analytics, 1978.

Guilford, J.P. Psychometric methods. New York: McGraw-Hill, 1954.

Hayes, W.L. Statistics for the social sciences. New York: Holt, Rinehart and Winston, 1973.

House, E.R. Evaluating with validity. Beverly Hills, Calif.: Sage Publications, 1980.

Irving, G.W., Hovinek, J.J., Walsh, D.H., & Chan, P.Y. ODA Pilot Study: Selection of an intensive graphics control device for continuous subjective functions applications. Santa Monica, Calif.: Integral Sciences Corporation, April 1976.

Kalenty, C.R., Lockwood, W.L., & Vissering, V.M.J. Experimental validation of an options selection matrix and investigation of other display formats as operational decision aids. Bethpage, New York: Grumman Aerospace Corporation, February 1977.

Keen, P.G.W. Computer-based decision aids: The evaluation problem. Sloan Management Review, 1975, 16(3), 11-29.

Kelly, G.A. The psychology of personal constructs, Vol. 1. New York: Norton, 1955.

Kenny, D.A. Correlation and causality. New York: John Wiley & Sons, 1979.

Kirk, R.E. Experimental designs: Procedures for the behavioral sciences. Belmont, Calif., Crooks/Cole Publishing Co., 1968.

Leal, A., Chen, K.K., Gardiner, P.C., & Frudy, A. Studies and application of adoptive decision aiding in anti-submarine warfare. Woodland Hills, Calif.: Perceptronics, April 1978.

Levin, H.M. Cost-effectiveness analysis in evaluation research. In M.Guttentag & E. Stuening (Eds.), Handbook of evaluation research, Vol. 2, Beverly Hills, Calif.: Sage, 1975.

Madden, E.G., & Siegel, A.I. Evaluations of operational decision aids:2. The emission control aid. Wayne, Pa.: Applied Psychological Services, April 1980.

Merkhofer, M.W., Robinson, B., & Korson, R.J. A computer based decision structuring process. Menlo Park, Calif.: SRI International, June 1979.

Milier, A.C.,III, Morris, P.A., Langford, R.L., Smallwood, R.D., & Gibbons, R.S. Analytic procedures for designing and evaluating decision aids. Menlo Park, Calif.: Applied Decision Analysis, April 1980.

Nagel, E. The structure of science. New York: Harcourt Brace Javanovich, 1961.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., & Bent, D.H. Statistical package for the social sciences. New York: McGraw-Hill, 1975.

Nunally, J.C. The study of change in evolution research: Principles concerning measurement, experimental design and analysis. In L.L. Struering, & M. Guttentag (Eds.), Handbook of evaluation research, Vol. 1. Beverly Hills, Calif.: Sage Publications, 1975.

Schechterman, M.D., & Walsh, D.H. Comparison of an operator aided optimization with iteractive manual optimization in a simulated tactical decision aiding task. Santa Monica, Calif.: Integrated Sciences, July 1980.

Shalit, B. Shalit perceptual organisation (sic) and reduction questionnaire (SPORQ). Stockhold: Forsvarets Forskningsanstalt, Part 1, 1978, Parts 2 and 3, 1979.

Siegel, A.I., & Federman, P.J. Validation of the DEI technique for large-scale display evaluation. Wayne, Pa.: Applied Psychological Services, 1967.

Siegel, A.I., Fischl, M.A., & Macpherson, D. The analytic profile system (APS) for evaluating visual displays. Human Factors, 1975, 17(3), 278-288.

Siegel, A.I., & Madden, E.G. Evaluations of operational decision aids: 1. The strike timing aid. Wayne, Pa.: Applied Psychological Services, January 1980.

Siegel, A.I., Miehle, W., & Federman, P. Short calculations for and validity of DEI technique. Wayne, Pa.: Applied Psychological Services, 1963.

Siegel, A.I., Miehle, W., & Federman, P. The DEI technique for evaluating equipment systems from the information transfer point of view. Human Factors, 1964, 279-286.

Siegel, A.I., & Wolf, J.J. Digital behavioral simulation--state-of-the-art and implications. Technical Report (Draft), Wayne, Pa.: Applied Psychological Services, March 1981.

Sinaiko, H.W. Operational decision aids: A program of applied research for naval command and control systems. ONR Report TR-5. Washington, D.C.: Manpower Research and Advisory Services, Smithsonian Institute, July 1976.

Stanley, J.C. Designing psychological experiments. In B.B. Wolman (Ed.), Handbook of general psychology. Englewood Cliffs, N.J.: Prentice Hall, 1973.

Walsh, D.H., & Schechterman, M.D. Experimental investigation of the usefulness of operator aided optimization in a simulated tactical decision aiding task. Santa Monica, Calif.: Integrated Sciences Corporation, January 1978.

Winer, B.J. Statistical principles in experimental designs. Second Edition. New York: McGraw-Hill, 1971.

Wolman, B.B. Concerning psychology and the philosophy of science. In B.B. Wolman (Ed.), Handbook of General Psychology. Englewood Cliffs, N.J.: Prentice Hall, 1973.

DISTRIBUTION LIST

Mr. J. Barber
HQS, Department of the Army
DAPE-MBR
Washington, D.C. 20310

Dr. Joseph Zeidner
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and
Systems Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director
U.S. Army Human Engineering Labs
Aberdeen Proving Ground, MD 21005

U. S. Army Medical R&D Command
ATTN: CPT Gerald P. Krueger
Ft. Detrick, MD 21701

ARI Field Unit-USAREUR
ATTN: Library
C/O ODCSPER
HQ USAREUR & 7th Army
APO New York 09403

U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D.C. 20332

Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Air University Library
Maxwell Air Force Base, AL 36112

Dr. Earl Alluisi
Chief Scientist
AFHRL/CCN
Brooks AFB, TX 78235

North East London Polytechnic
The Charles Myers Library
Livingstone Road
Stratford
London E15 2LJ ENGLAND

Professor Dr. Carl Graf Hoyos
Institute for Psychology
Technical University
8000 Munich
Arcisstr 21
FEDERAL REPUBLIC OF GERMANY

Dr. Kenneth Gardner
Applied Psychology Unit
Admiralty Marine Technology
Establishment
Teddington, Middlesex TW11 OLN
ENGLAND

Director, Human Factors Wing
Defense & Civil Institute of
Environmental Medicine
Post Office Box 2000
Downsview, Ontario M3M 3B9
CANADA

Dr. Rex Brown
Decision Science Consortium
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

Dr. A. D. Baddeley
Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF
ENGLAND

Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22314 (12 cys)

Dr. Craig Fields
Director, Cybernetics Technology
Office
Defense Advanced Research Projects
Agency
1400 Wilson Blvd.
Arlington, VA 22209

Dr. Judith Daly
Cybernetics Technology Office
Defense Advanced Research Projects
Agency
1400 Wilson Blvd.
Arlington, VA 22209

Dr. Lloyd Hitchcock
Federal Aviation Administration
ACT 200
Atlantic City Airport, NJ 08405

Dr. M. Montemerlo
Human Factors & Simulation
Technology, RTE-6
NASA HQS
Washington, D.C. 20546

Professor Douglas E. Hunter
Defense Intelligence School
Washington, D.C. 20374

Dr. Robert R. Mackie
Human Factors Research, Inc.
5775 Dawson Avenue
Goleta, CA 93017

Dr. Miley Merkhofer
Stanford Research Institute
Decision Analysis Group
Menlo Park, CA 94025

Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, VA 22202

Professor Judea Pearl
Engineering Systems Department
University of California-Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90024

Professor Howard Raiffa
Graduate School of Business
Administration
Harvard University
Soldiers Field Road
Boston, MA 02163

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Mr. Joseph G. Wohl
Alphatech, Inc.
3 New England Executive Park
Burlington, MA 01803

Dr. Robert T. Hennessy
NAS-National Research Council
JH #819
2101 Constitution Ave., N.W.
Washington, D.C. 20418

Dr. M. G. Samet
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. Robert Williges
Human Factors Laboratory
Virginia Polytechnical Institute
and State University
130 Whittemore Hall
Blacksburg, VA 24061

Dr. Alphonse Chapanis
Department of Psychology
The Johns Hopkins University
Charles and 34th Streets
Baltimore, MD 21218

Dr. Meredith P. Crawford
American Psychological Association
Office of Educational Affairs
1200 17th Street, N.W.
Washington, D.C. 20036

Dr. Ward Edwards
Director, Social Science Research
Institute
University of Southern California
Los Angeles, CA 90007

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, OK 73069

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Room 201
Boulder, CO 80309

Journal Supplement Abstract Service
American Psychological Association
1200 17th Street, N.W.
Washington, D.C. 20036 (3 cys)

Mr. Edward M. Connelly
Performance Measurement Assoc., Inc.
410 Pine Street, S.E. Suite 300
Vienna, VA 22180

Dr. Richard W. Pew
Information Sciences Division
Bold Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

Mr. Tim Gilbert
The MITRE Corporation
1820 Dolly Madison Blvd.
McLean, VA 22102

Dr. Baruch Fischhoff
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Andrew P. Sage
University of Virginia
School of Engineering and Applied
Science
Charlottesville, VA 22901

CDR Paul R. Chatelier
Office of the Deputy Under Secretary of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D.C. 20301

Dr. Stuart Starr
Office of the Assistant Secretary of Defense (C3I)
Pentagon
Washington, D.C. 20301

Director
Engineering Psychology Programs
Code 455
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217 (5 cys)

Director
Communication & Computer Technology
Code 240
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director
Tactical Development & Evaluation Support
Code 230
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director
Manpower, Personnel and Training
Code 270
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director
Operations Research Programs
Code 434
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director
Information Systems Program
Code 437
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Code 430B
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Special Assistant for the Marine Corps Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Commanding Officer
ONR Eastern/Central Regional Office
ATTN: Dr. J. Lester
Building 114, Section D
666 Summer Street
Boston, MA 02210

Commanding Officer
ONR Branch Office
ATTN: Dr. C. Davis
536 South Clark Street
Chicago, IL 60605

Dr. Leonard Adelman
Decisions and Designs, Inc.
8400 Westpark Drive, Suite 600
P.O. Box 907
McLean, VA 22101

Commanding Officer
ONR Western Regional Office
ATTN: Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106

Commanding Officer
ONR Western Regional Office
ATTN: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
American Embassy, Room A-407
APO San Francisco, CA 96503

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375 (6 cys)

Dr. Michael Melich
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, D.C. 20375

Dr. Robert G. Smith
Office of the Chief of Naval Operations, OP987H
Personnel Logistics Plans
Washington, D.C. 20350

Dr. Jerry C. Lamb
Combat Control Systems
Naval Underwater Systems Center
Newport, RI 02840

Naval Training Equipment Center
ATTN: Technical Library
Orlando, FL 32813

Human Factors Department
Code N215
Naval Training Equipment Center
Orlando, FL 32813

Dr. Albert Colella
Combat Control Systems
Naval Underwater Systems Center
Newport, RI 02840

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Mr. Warren Lewis
Human Engineering Branch
Code 8231
Naval Ocean Systems Center
San Diego, CA 92152

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D.C. 20380

Commanding Officer
MCTSSA
Marine Corps Base
Camp Pendleton, CA 92055

Mr. Wayne Zachary
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

Chief, C^3 Division
Development Center
MCDEC
Quantico, VA 22134

Mr. Arnold Rubinstein
Naval Material Command
NAVMAT 0722 - Rm. 508
800 North Quincy Street
Arlington, VA 22217

Commander
Naval Air Systems Command
Human Factors Programs
NAVAIR 340F
Washington, D.C. 20361

Mr. Phillip Andrews
Naval Sea Systems Command
NAVSEA 0341
Washington, D.C. 20362

Commander
Naval Electronics Systems Command
Human Factors Engineering Branch
Code 4701
Washington, D.C. 20360

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Dr. James McGrath, Code 302
Navy Personnel Research and Development Center
San Diego, CA 92152

Navy Personnel Research and Development Center
Planning & Appraisal
Code 04
San Diego, CA 92152

CDR Norman Lane
Code 6021
Naval Air Development Center
Warminster, PA 18974

Dr. Julie Hopson
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Jeffrey Grossman
Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

Human Factors Engineering Branch
Code 1226
Pacific Missile Test Center
Point Mugu, CA 93042

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

CDR W. Moroney
Code 55MP
Naval Postgraduate School
Monterey, CA 93940

Mr. Merlin Malehorn
Office of the Chief of Naval Operations (OP-115)
Washington, D.C. 20350

Dr. Thomas Hammell
Eclectech Associates, Inc.
P.O. Box 178
North Stonington, CT 06359

END

DATE
FILMED
10-81

DTIC

FRONT END A

SYST

- WH
- HO
- HO
- TIM
- RE
- AS
- RIS
- TE
- OP

DESI
DEVI

- PRE

Nagel, E. The structure of science. New York: Harcourt Brace Javanovich, 1961.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., & Bent, D.H. Statistical package for the social sciences. New York: McGraw-Hill, 1975.

Wolman, B.B. Concerning psychology and the philosophy of science. In B.B. Wolman (Ed.), Handbook of General Psychology. Englewood Cliffs, N.J.: Prentice Hall, 1973.

Defense & Civil Institute of
Environmental Medicine
Post Office Box 2000
Downsview, Ontario M3M 3B9
CANADA

Dr. Rex Brown
Decision Science Consortium
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

1201 Oak Street
Eugene, OR 97401

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Mr. Joseph G. Wohl
Alphatech, Inc.
3 New England Executive Park
Burlington, MA 01803

1201 Oak Street
Eugene, OR 97401

Dr. Andrew P. Sage
University of Virginia
School of Engineering and Applied
Science
Charlottesville, VA 22901

Boston, MA 02210

Commanding Officer
ONR Branch Office
ATTN: Dr. C. Davis
536 South Clark Street
Chicago, IL 60605

Dr. Leonard Adelman
Decisions and Designs, Inc.
6400 Weatpark Drive, Suite 600
P.O. Box 907
McLean, VA 22101

Washington, D.C. 20360

Commanding Officer
MCTSSA
Marine Corps Base
Camp Pendleton, CA 62055

Mr. Wayne Zachary
Analytics, inc,
2500 Maryland Road
Willow Grove, PA 19090

Monterey, CA 93940

Mr. Merlin Malehorn
Office of the Chief of Naval
Operations (OP-115)
Washington, D.C. 20350

Dr. Thomas Hammell
Eclectech Associates, Inc.
P.O. Box 179
North Stonington, CT 06359