Technical Report 611

Issues in the Design and Evaluation of Decision–Analytic Aids

Leonard Adelman, Michael L. Donnell,
John F. Patterson, and Jonathan J. Weiss
Decisions and Designs, Incorporated

Battlefield Information Systems Technical Area
Systems Research Laboratory

U. S. Army
Research Institute for the Behavioral and Social Sciences

January 1984

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

L. NEALE COSBY
Colonel, IN
Commander

EDGAR M. JOHNSON
Technical Director

Research accomplished under contract for
the Department of the Army

Decisions and Designs, Incorporated

Technical review by

Mary Jo Hall
Ruth H. Phelps

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POT, 5001 Eisenhower Avenue, Alexandria, Virginia 22333.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report 611	2. GOVT ACCESSION NO.: AD-A148 313
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): ISSUES IN THE DESIGN AND EVALUATION OF DECISION-ANALYTIC AIDS	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER: --
7. AUTHOR(s): Leonard Adelman, John F. Patterson, Michael L. Donnell, Jonathan J. Weiss	8. CONTRACT OR GRANT NUMBER(s): MDA903-80-C-0194
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Decisions and Designs, Inc. Suite 600, 8400 Westpark Drive, P.O. Box 907 McLean, VA 22101	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2Q263739A793
11. CONTROLLING OFFICE NAME AND ADDRESS: US Army Research Institute for the Behavioral and Social Sciences. 5001 Eisenhower Avenue, Alexandria, VA 22333-5600	12. REPORT DATE: January 1984
	13. NUMBER OF PAGES: 48
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: --

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

--

18. SUPPLEMENTARY NOTES

This report is the result of a colloquium held at Decisions and Designs, Inc., on April 21, 1980. Dr. Stanley Halpin, Dr. Ruth Phelps, Dr. William Deeds, and Dr. Uldi Shvern represented the Army Research Institute and Dr. Clinton Kelly, III, Dr. Michael Donnell, Dr. Jonathan Weiss, (Continued)

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

Decision theory
Decision analysis
Decision making
Decision aids
Mathematical models

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

Decision analysis has emerged as a highly valuable technology for allowing decision makers to formulate important problems in a logical framework, incorporating factual as well as judgmental information to arrive at consistent, realistic solutions. Computers have served well as aids to calculation, display, editing, and memory functions. On the basis of previous success, organizations are beginning to develop computer-based decision-analytic aids with stand-alone capabilities for routine use by internal analysts and decision (Continued)

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

UNCLASSIFIED

 SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Item 18 (Continued)

and Dr. Leonard Adelman represented DDI. Ruth Phelps and Jo Hall made major contributions to the colloquium organization and content and to manuscript preparation. This report was sponsored in conjunction with Defense Advanced Research Projects Agency (DARPA) Cybernetics Technology Division, Defense Sciences Office, 1400 Wilson Blvd., Arlington, VA 22209.

Item 20 (Continued)

makers without outside consultation. Decision-analytic aids include different types of multi-attribute utility assessment models and traditional decision-theoretic tree models requiring probability and utility assessments. Although some stand-alone decision-analytic aids have been quite successful, others have not been utilized by their prospective users. The purpose of this report is to provide guidelines for the effective design, implementation, and evaluation of such decision aids.

A framework for considering issues relevant to the design and evaluation of decision-analytic aids is presented in the introduction. This framework identifies three interfaces essential for the effective integration of decision aids into organizations. The first interface is between the decision aid and the user. Here, the issue is the extent to which characteristics of the aid facilitate or hinder its usability. The second interface is between the user (and decision aid) and the larger decision-making organization. Here, the question is to what extent the decision aid facilitates the decision-making processes of the organization. The third interface is between the decision-making organization and the environment. Here, the issue is whether the aid improves the quality of the organization's decision making. The sections of this report sequentially consider the issues at each of the three interfaces and provide guidelines for effectively addressing those issues.

The authors realize that these guidelines will not answer all the questions of potential developers and users of decision aids, for the development of such aids is less than one decade old. This report does identify, however, those issues in development and evaluation that have arisen in the work of decision analysts over the past few years. Such information should assist developers in integrating decision aids into their organizations and, in turn, result in improved organizational decision making.

Technical Report 611

Issues in the Design and Evaluation of Decision-Analytic Aids

Leonard Adelman, Michael L. Donnell,
John F. Patterson, and Jonathan J. Weiss
Decisions and Designs, Incorporated

Ruth H. Phelps, Contracting Officer's Representative

Submitted by
Harold Martinek, Acting Chief
Battlefield Information Systems Technical Area

Approved as technically adequate
and submitted for publication by
Jerrold M. Levine, Director
Systems Research Laboratory

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

January 1984

Army Project Number
2Q263739A793

Human Factors in Training &
Operational Effectiveness

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The Battlefield Information Systems Technical Area is concerned with the demands of the future battlefields for increased human-machine capacity to acquire, transmit, process, disseminate, and utilize information. Research is focused on interface problems and interactions within command, control, and intelligence centers and is concerned with such areas as tactical symbology, user-oriented systems, information management, staff operations and procedures, and sensor systems integration and utilization.

One area of special interest is the development of procedures to support and enhance the decision-making process within command, control, and intelligence centers. The current effort summarizes guidelines for the effective design, implementation, and evaluation of decision-analytic aids. Also presented in the framework from which these guidelines were developed. This framework identifies three interfaces essential to the integration of decision aids into organizations. This report, therefore, should assist developers in integrating decision aids into user organizations, resulting in improved decision making.

Research in decision aiding is conducted as an in-house effort with additional support from contracting organizations that are selected for their unique contributions to this area. This effort is responsive to the requirements of Army Project 2Q263739A793 and was managed through the Cybernetics Technology Office of DARPA.

EDGAR M. JOHNSON
Technical Director

ISSUES IN THE DESIGN AND EVALUATION OF DECISION-ANALYTIC AIDS

EXECUTIVE SUMMARY

Requirement:

To formulate and demonstrate guidelines for developers to use in the design, implementation, and evaluation of decision aids.

Approach:

The guidelines were developed from a framework based on three interfaces essential to the integration of decision aids in organizations. These interfaces refer to the contact points between the user and the aid, the aid and the decision-making organization, and the organization and its associated environment.

Product:

Design issues arising from the interfaces are addressed. Special attention is devoted to (1) the immediate behavioral effects of decision aids and the engineering of aid software and hardware so as to minimize adverse consequences and (2) the importance of user involvement in aid design to ensure the understanding and commitment necessary to adopt a different decision-making approach, as well as to tailor the aid to the users' particular needs. Evaluation issues are considered, with a focus on four major areas: (1) factors at each of the three interfaces that make aids effective, (2) evaluation settings (their similarity to operational settings, the amount of control they provide, and their costs), (3) methods for obtaining measures of effectiveness, and (4) control conditions required for adequate evaluation.

Utilization:

This report should provide decision aid designers with guidelines and issues of concern in the cycle of development, from conceptualization and implementation to evaluation and revision of the aid. The information should assist developers in integrating decision aids into the intended organizations and result in improved organizational decision making.

ISSUES IN THE DESIGN AND EVALUATION OF DECISION-ANALYTIC AIDS

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

ISSUES IN THE DESIGN AND EVALUATION OF DECISION-ANALYTIC AIDS

INTRODUCTION

During recent years, decision analysis has emerged as a highly valuable technology for allowing decision makers to formulate important problems in a logical framework, incorporating factual as well as judgmental information to arrive at consistent, realistic solutions. Computers have served well as aids to calculation, display, editing, and memory functions. On the basis of previous success, organizations are beginning to develop computer-based decision-analytic aids with stand-alone capabilities for routine use by internal analysts and decision makers without outside consultation. Although some stand-alone decision aids have been quite successful, others have not been utilized by their prospective users. The purpose of this report is to provide guidelines for the effective design, implementation, and evaluation of such decision aids.

Throughout the following sections, the term <u>decision aid</u> or <u>decision-analytic aid</u> refers to a computer that has been programed to assist in formulating and exercising decision-theoretic models. These include different types of multi-attribute utility assessment models and traditional decision-theoretic tree models requiring probability and utility assessments. Aids may take on a variety of forms, from the simplest of clerical devices implemented on micro- or mini-computers (special-purpose routines to perform calculations and to display or store results) to the most sophisticated, state-of-the-art, large-scale computer implementations (general-purpose aids that help the user structure a wide variety of problems, search through large data bases, and perform complex analyses). But whatever their role, decision aids are designed to provide one or both of the following primary benefits:

- improved decision quality--the assurance that a decision is logically based on a consistent, explicit, and realistic set of assumptions; and
- lower decision costs--a saving in some critical resource (time, money, personnel, etc.), compared with the unaided decision process.

To be sure, additional benefits may accrue: The decision maker may develop greater understanding of the overall problem area, or may find computer-aided solutions easier to implement, but unless a device either improves or facilitates decision making, it cannot properly be termed a decision aid.

Figure 1 is a pictorial representation of the framework for considering issues relevant to the design and evaluation of decision aids. These issues arise at three interfaces represented in the figure. The first interface is between the decision aid and the user. Here, the issue is the extent to which characteristics of the aid facilitate or hinder its usability. The second interface is between the user (and decision aid) and the larger decision-making organization. Here, the question is to what extent the decision aid facilitates the decision-making processes of the organization. The third interface is between the decision-making organization and the environment. Here, the issue is whether the aid improves the quality of the organization's decision making. The sections of this report sequentially address the issues at each of the three interfaces.

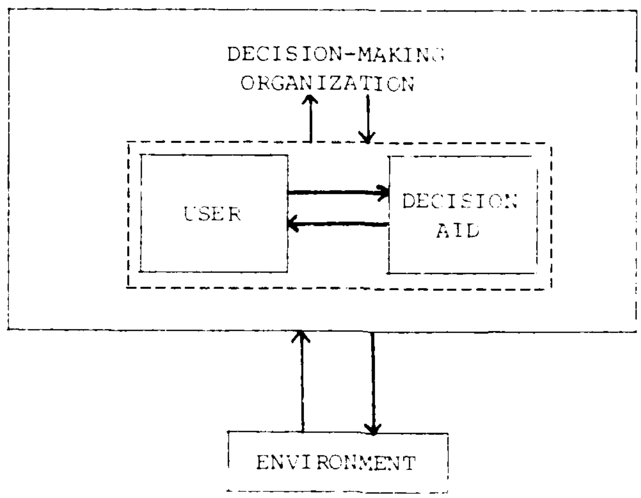


Figure 1. Framework for considering issues relevant to the design and evaluation of decision aids.

The section on behavioral engineering addresses the first interface, starting with the following premise about human behavior: Long-range benefits, especially uncertain and unattributable benefits, are generally underweighted when compared with immediately observable effects. Thus, even though a decision aid may be objectively acknowledged as a worthwhile investment of time and effort, the user's immediate behavior may be dominated by short-range perceptions of increased workload and by feelings of boredom, impatience, frustration, or embarrassment that stem not from the process itself, but from its implementation. As a result, specific efforts must be made to analyze the immediate behavioral effects of decision aids and to engineer these aids so as to minimize or reverse their adverse consequences. Some of the major behavioral problems typical of current decision aids are discussed, and possible solutions to these problems are offered.

The section on involving users in the development of decision aids addresses the second interface. The thesis of this section is that decision aids will seldom achieve stand-alone status unless eventual users (both hands-on users and decision makers) are involved in their development. Solving behavioral problems at the first interface is necessary, but not sufficient, for aid implementation. User involvement in aid design is also essential for implementation, for this involvement develops the understanding and commitment necessary for implementing a different decision-making approach and tailors aid characteristics to the users' needs within their organizational context. Two sources of support for this position are presented: (1) the recent systematic evaluation of a decision aid developed for the Operations Directorate of the Joint Chiefs of Staff and (2) the history of model implementation in the field of operations research and management science. Although the need for user involvement appears obvious, these sources indicated that user involvement is often neglected in the development of analytical decision aids, frequently resulting in unsuccessful implementation efforts.

The final section focuses not on the design and implementation of decision aids, but on decision aid evaluation. In particular, this section focuses on four major aspects of decision aid evaluation. The first aspect concerns the factors that make a decision aid effective. These factors determine which general measures of effectiveness are employed to evaluate the aid in relation to previously specified objectives. The second aspect concerns the settings for decision aid evaluations. Evaluation settings differ in their similarity to the expected operational setting, the amount of control they provide, and their costs; such differences affect the extent to which certain measures of effectiveness can be collected and analyzed during the evaluation. The third aspect concerns the different methods for obtaining measures of effectiveness. The fourth aspect addresses the problem of developing adequate control (or contrasting) conditions for the effective evaluation of decision aids. Each of the four aspects addresses issues at each of the three interfaces.

It is hoped that this report will provide effective guidelines for the design and evaluation of decision aids. The authors realize that these guidelines will not answer all the questions of potential developers and users of decision aids, for the development of such aids is less than one decade old. The report does identify, however, those issues in development and evaluation that have arisen in the work of decision analysts over the past few years. Such information should assist developers in integrating decision aids into their organizations and, in turn, result in improved organizational decision making.

BEHAVIORAL ENGINEERING: CHALLENGE TO DECISION AID DESIGN

The historical failure of most decision aids to generate user enthusiasm can be largely explained in terms of a well-known principle of behavioral psychology: Immediate, certain, directly observable effects have high impacts on behavior, whereas delay, uncertainty, and indirectness of results can reduce the perceived impact of rewards (benefits) and punishments (costs). If the behavior in question is decision aid usage, the costs consist of immediate expenditures of time, effort, and attention in a stressful, time-constrained situation. The benefits, in contrast, are deferred until the analysis is complete and the decision implemented; even then, uncertain future events may make a rational decision look bad. Furthermore, many decisions have their primary effect on other members of the organization, so that only indirect feedback reaches the decision maker. Thus, the immediate process-related cost factors dominate the behavioral environment of decision aiding, and the results-related benefits have less impact than they should.

This effect is compounded by the fact that a successful decision analysis is not a single event, but a complex, prolonged sequence of behavior that requires continuously high levels of motivation and attention. There are many opportunities for the user to become bored, confused, or discouraged, and any one of these may cause the user to terminate the analysis or, even worse, may lead to a half-hearted attempt at analysis that not only compounds the behavioral problems, but also increases the risk of unnoticed analytic errors. It is not sufficient that the long-range benefits of using a decision aid outweigh its operating costs; at every point in the process, the user must perceive immediate rewards to be motivated to continue the analysis.

This principle is illustrated by the role of a professional decision analyst in a clinical decision analysis project. Even though the client has already made a conscious decision to employ the decision analyst's methodology, the analyst must be prepared to deal with occasional episodes of confusion, fatigue, boredom, impatience, and discouragement. Without proper treatment, these problems may delay the analysis, degrade the quality of the outcome, or reduce the client's confidence in the process. Therefore, the analyst must be more than technically proficient; the analyst must be sensitive to the client's mental and emotional state and must have the flexibility to adapt the direction and timing of the analysis accordingly. Instead of proceeding in a linear fashion through an analysis, the analyst may need to stop and review results, explain the procedures more carefully, repeat a portion of the analysis, call for a break, or simply administer some reassurance and encouragement to continue. In short, client motivation and attention are as basic to the function of the decision analyst as is technical performance, and perhaps even more important, as it is easier to recover from a technical error than from a motivational one.

It is a major challenge for the professionally trained decision analyst to maintain motivation and attention consistently, but it is an even greater challenge for the decision aid to accomplish the same task because the aid is not able to observe the user's emotional and mental state, cannot infer and adapt to the user's unique personal characteristics, and has no personal credibility in a leadership role. In order to maintain motivation and attention, the computerized decision aid must compensate for these shortcomings by capitalizing on its strengths--speed, precision, memory capacity, and the ability to generate neat and effective output displays--to minimize the perceived costs of usage, and to offset those costs with even greater immediate benefits.

During the past decade, two microprocessor-based technologies have had dramatically different fates. Commercially produced video games (e.g., "Pong," "Tank War," and "Space Invaders") have achieved rapid and widespread acceptance in the entertainment market, while computerized decision aids have generated very little enthusiasm among their intended users. The play-versus-work distinction by itself does not explain the divergent fates of these two young technologies; but in their efforts to capture a highly competitive market, the designers of video games have incorporated features of behavioral engineering that decision aid designers have largely ignored. Thus, the success of the video games can be regarded as a challenge to the decision aid designers, suggesting new directions to guide further development, while offering optimistic evidence for eventual success.

The major contrast between the current generation of decision aids and the far more successful video games is that the former depend in large part on the user's ability to endure unpleasant work in order to achieve a worthwhile long-range result, whereas the latter focus on more immediate features: simplicity of operation, speed of response, direct sensory feedback, and frequent rewards. This section suggests several ways in which decision aids could be designed to take advantage of the behavioral principles that favor the latter approach.

Training, Initial Access, and Startup

Probably the most critical hurdle is to get the user to sit down in front of the machine, turn it on, and start working. At this initial stage, benefits are the most remote, inertia is strongest, the potential user has nothing invested in the process, and a variety of easier alternatives present themselves. Before addressing any substantive issues, the potential user must determine whether and how to use the aid, learn or refresh the necessary procedural knowledge, obtain appropriate access to the aid, and follow the prescribed startup procedures. Although this phase is intrinsically an investment of sorts, it is essential to minimize the effort and stress involved and to provide as much assistance and encouragement as possible. In particular, the following goals can help to overcome the initial block to usage.

Clear Product Identification. A decision aid should come with a clear label designating the aid's intended use, the time and training it requires, its products, and a reference to a further source of information. If possible, a hot-line telephone should be available for instant assistance; failing this, at least one person or office might assume primary responsibility for providing users with information.

Aesthetic Package Design. The decision aid should appear functional and convenient to use. It should appear to be an efficient, robust, and task-oriented machine, solidly constructed without much unnecessary clutter. The controls and displays should appear as simple as possible to avoid intimidating the potential user, and all features should be clearly labeled. Within these constraints, further efforts should focus on the visual appeal of the machine itself, particularly the appeal of the area the user will be occupying.

Training and Documentation. Although a detailed system description and a technical reference document should be available for reference or for higher level training, the ordinary user should need a minimum of instruction and documentation. Ideally, the system should incorporate a self-contained, optional tutorial rather than an off-line training manual; for ordinary use, the startup instructions should be few and simple enough to appear on a panel attached to the machine.

Access and Availability. The physical setup of the decision aid and the startup routine's software should be designed to minimize startup effort. A single switch or control should turn power on (if necessary), and an automatic loading routine should initiate the startup procedure without further commands. Any administrative procedures such as accounting, user identification, and authorization should, if possible, be taken care of at the site of the aid, with specially designed routines to minimize the amount of paper work required. For example, instead of filling out an authorization form, a user might simply type in a name or identification code, allow the machine to prepare the necessary documents, and sign the completed form.

User Participation in Customized Design. When an aid is developed with a particular user population in mind, the opinions of representatives from the user population should be solicited and then incorporated in the aid's design. Apart from the direct effect of customizing the product to meet unique needs and to conform as much as possible to current procedures and conventions, the participation of the user population will facilitate training and help smooth the transition process. Most important, by involving users in the design of the system, the designers can overcome the users' natural tendency to oppose the imposition of an unfamiliar process by outside forces, and instead foster a feeling of personal investment in the success of the aid. The next section of this report considers in detail the importance of user involvement for successful implementation.

Ease and Effectiveness of Operation

Once the user has decided to employ a decision aid, success depends on the aid's ability to help the user through the procedures without disruption due to lapses in attention or motivation. In the on-going dialogue, the aid's outputs should be clear and useful, and the required user inputs should impose a minimum of strain on the user. To achieve these goals, a successful decision aid should incorporate simple control mechanisms, continuous sensorimotor feedback, effective design of output formats, and human-engineered procedural requirements.

Simple Controls. One of the most frequent complaints of potential decision aid users (particularly high-level decision makers) is that they are required to type all their inputs on a keyboard. Although an experienced typist or computer user might feel comfortable with a keyboard-based input method, a less experienced typist might experience impatience because of limited typing speed, frustration because of high error rates, embarrassment in front of onlookers, and, in some cases, a feeling that the use of the aid is primarily a clerical chore rather than a more challenging analytic task. Furthermore, while concentrating on avoiding typographical or spelling errors, a decision maker is distracted from the substantive task at hand. When time and organizational pressures severely constrain the decision maker, these effects may increase so significantly that the aid is abandoned.

This problem could be avoided by using a typist or a specially trained operator to enter the user's orally expressed responses, but such an approach would increase operating costs and personnel requirements, would make the aid less accessible, and, most important, would eliminate the direct private and personal involvement of the decision maker. Therefore, the use of a clerical assistant is unacceptable in most cases.

Fortunately, some simple alternate approaches can eliminate the need for most keyboard inputs. These approaches may range from the simple buttons and levers that control commercial computer games to futuristic devices such as speech recognizers, visual pattern detectors, and pressure-sensitive display screens. Although further human-engineering efforts will be needed to determine the most effective functions, sizes, positions, and configurations, these simplified input devices can already produce big improvements in user engineering. As more advanced technology becomes available, it will become even easier for the untrained user to enter information and to issue commands quickly, painlessly, and reliably.

Continual Sensorimotor Feedback. When two people interact in normal conversation, the information transmitted is more than just the words that are being uttered. The "carrier wave" that makes a conversation seem natural is an almost unnoticed background of eye motions, nonverbal noises, postural adjustments, and facial movements, all of which transmit such procedural messages as "I am still listening," "I have something to say," or "I am not sure I understand what you mean." Without this continuous nonverbal interaction, the conversation would become awkward and the participants would feel uncomfortable. Occasional long pauses would occur, making the participants unsure about whether the dialogue should continue. At other times, two or more parties would attempt to speak at once, inducing not only confusion but a certain amount of social friction as well. The nonverbal component of a face-to-face conversation helps to avoid these problems, smoothing the social content and assuring efficient communication.

The same observation applies to the interaction between the decision aid and its user; the aid must continually assure the user that procedures are in normal working order, that inputs are being properly received, and that output responses are on schedule. To provide continuous or frequent indication that all functions are working properly, for example, the aid may include a clock whose display changes every second. To provide immediate sensorimotor feedback by acknowledging every input without perceptible delay, the aid may use a visual signal, a simple tone, or an echo of the user's input. This feature is especially necessary when time is of the essence and the user is preoccupied with the overall process; even a fraction of a second without response can arouse either impatience (in the case of experienced users who have previously been led to expect instantaneous responses) or fear of computer malfunction (in the case of inexperienced users who simply do not know whether a delay is normal). Whenever an especially long delay (more than 30 seconds) is absolutely necessary, the aid should not only acknowledge the input that initiated the long operation, but also provide an estimate of the time needed to complete it and an option to cancel a very long operation if the user so wishes. During the delay, music or graphical displays can be used to maintain contact and hold the user's attention.

Effective Output Formats. The perceived value of an aid's outputs will depend not only on the contents of the displays and printouts, but also on their format and style. A well-designed output is easily read. It should direct the user's attention to the proper information, facilitate the user's ability to focus on selected items, and all the while remain aesthetically attractive. The following ideas illustrate possible ways of achieving these goals.

- Replace textual alphanumeric outputs with such formats as pictorial symbols, photographic images, graphs, and maps.

- Use prerecorded or synthesized speech to present passages of text; to provide procedural guidance; to annotate graphs, charts, and maps; and to add emphasis to important alphanumeric display messages.

- Use unique nonverbal sounds (notes, chords, noises) to attract the user's attention whenever something unusual demands an especially high level of alertness.

- Use motion and color to direct the user's visual attention to specific segments of the displays.

- Use graphics and music to enhance the aid's aesthetic appeal and its overall image of quality.

Human Factors Engineering. Even though every individual task associated with the decision aid may be easily within the user's grasp, the combination of tasks to be accomplished in a short time span may exceed the user's limits. An approach using human factors engineering can identify those points where overloads of this type are likely and then indicate ways in which the redesign of some portion of the aid can reduce overload. For example, if the user must attend to a number of stimuli simultaneously, it may be possible to present them through different sensory channels or to present them sequentially to avoid confusion. Altering a display configuration or an input device may have a significant impact on the amount of strain imposed, and improving timing and sequencing may allow users to work more effectively without reducing the actual task requirements. Finally, designing specific color-coding schemes, symbols, and auditory cues to correspond with the user's "natural" expectations may relieve some of the effort involved in interpreting outputs.

Minimal Requirements for Technical Knowledge. Although it may be impossible to make an aid's decision-analytic techniques completely transparent to the user, every effort should be made to minimize the requirements for specialized methodological training. In no case should a technical decision-analytic term be used without explanation of its specialized meaning; if at all possible, technical jargon that might intimidate, confuse, or alienate the user should be avoided altogether. If analytic methods must be referred to, it might be preferable to invent new terms rather than risk the confusion that might arise from using ambiguous decision-analytic terms such as utility, attribute, risk, weight, and option. If the goal is to communicate with a naive user, there is little reason to insist on traditionally accepted terms. Of course, if the analysis can be conducted at the level of direct judgments (such as binary choices), keeping the decision-analytic implications of the user's responses internal to the aid's program, so much the better.

Intermediate Reinforcement to Increase Attention and Motivation

The preceding discussion has considered ways to attract the user to the aid and to simplify overall operation of the aid. However, because the user is a human being with a limited attention span and other responsibilities competing for time and attention, the aid must do more than just smooth the path toward the ultimate goal. In addition, the aid must help the user to follow that path without losing sight of the goal or being distracted along the way. Because the successful use of a decision aid requires a long chain of behaviors, many of which will be quite unexciting, the user will face boredom, fatigue, and impatience at times when the goal still appears quite remote. Insofar as possible, the aid should counter those effects by--

- reminding the user of the ultimate goal,
- providing milestones and progress reports along the way,
- rewarding the user for completing intermediate goals,
- timing and sequencing tasks to avoid boredom, and
- reinforcing the user for maintaining a high level of attention.

Goal Focusing. One useful way to keep the decision maker aimed toward the desired destination is to provide a sort of road map in the form of a milestone chart. This not only reminds the user of the ultimate goal, but provides a set of more modest subgoals for the user to complete. As each subgoal is reached, it can be represented on a progress chart, thereby rewarding the decision maker while pointing toward the next task. The completion of a subgoal might be a good occasion for a break or for a review of the partial results available. These results, in the form of hard-copy charts, graphs, tables, and text, can act as a further reward by providing the user with valuable information and tangible evidence of work completed.

Timing, Sequencing, and Variation. No matter how easy or enjoyable a task, it will eventually lose the user's interest if it is too prolonged or repeated too often. Satiation with task rewards, habituation to the visual and aural stimuli presented, and general fatigue will increase until motivation drops, attention lags, and error rates rise.

By dividing the overall task into shorter segments and varying successive tasks (e.g., using different sensory modalities, different display colors, different muscular movements), the decision aid can keep the user more attentive and better motivated. Human engineering can achieve the right balance between the attentional benefits to be gained from shorter tasks and the possible confusion and delay involved in switching tasks too frequently. Further study might identify groups of complementary tasks that could be effectively organized into a recurring cycle, to provide the necessary variety without unnecessary shifts in attention. Ideally, the transition from task to task should be significant and frequent enough to prevent boredom and fatigue, yet smooth and logical enough to maintain continuity.

Direct Reinforcement for Task-Oriented Behavior. Although some motivation may result simply from attaining intermediate goals, this source of reinforcement can be simply and directly augmented by providing more direct rewards as well. At the end of a given task sequence, the user might be given the opportunity to clear his or her mind by engaging in some sort of recreational activity for a limited time. The aid might, for example, provide a choice between a video game, a passage of recorded music, a selection of puzzles or jokes, and a display of computer art. In order to control the amount of time spent on such extraneous pursuits, access to the reward activities might be programed to occur only at random times, contingent upon successful completion of subgoals. Behavioral research has shown that random reinforcement of this sort is often far more efficient at maintaining effort than regular schedules of reinforcement with the same overall frequency of reward.

A further way of reinforcing attention during the performance of a task (without distracting the user from the task itself) might be to measure the user's response times, providing feedback in the form of occasional performance reviews (to be presented at the task's completion) and bonus rewards for good performance. For example, a numerical alertness score based on the user's speed of response might be combined with an error rate, or other behavioral measure, to determine the likelihood of a reward at the end of each task segment. Auditory feedback might be useful for this function in much the same way that the bells, clicks, and various electronic sounds reinforce the users of video games and pinball machines without impairing their attention. Because preferences and needs in this area may vary widely from user to user, the ability to involve users in the initial design (discussed in the next section in this report) or to adapt procedures to individual needs (discussed later in this section) will be especially important.

Final Products

Once the user has completed the analysis, the decision aid should provide as much reinforcement as possible in order to make its use more attractive in the future, while continuing to offer whatever support is available to translate the results of the analysis into action. The reinforcement might include the following:

- hard copies of tables, charts, graphs, and other materials that might be useful to brief others on the outcome of the analysis (or to input into some higher order decision process);
- a preformatted report that presents the analytic results and rationale in a readable format, along with supporting documentation on the analytic methods used and the conclusions reached; or, perhaps,
- a printed or videotaped protocol of the entire session, including a visual record of what has appeared on the display screen and an audio or textual record of inputs and verbal outputs.

Further assistance might take the form of follow-on analysis routines. For example, once the user has selected an overall course of action, the aid might offer an option to help construct a more detailed implementation plan. If sensitivity analyses indicate the need for better data on some critical topics, a value of-information analysis might help determine which data to collect and how extensive an effort is needed. Similarly, if a short-range decision has been made, the aid might provide some help toward integrating it with the related mid-to-long-range considerations.

Finally, whatever reinforcements were available upon completion of the subgoals ought to be presented (with certainty and in greater quantity) when all the session's work has been completed. Summary feedback on the user's behavioral data (response times, error rates, etc.) might be useful for the user's own benefit, although care should be taken to preserve the user's confidence in the privacy of this information. Unless time is extremely short, the user should be permitted to enjoy the recreational rewards and the satisfaction of having finished the complete analysis. Most important, the aid should acknowledge the user's hard work and elicit any comments, suggestions, or questions that might help to improve future versions of the aid.

Adapting Procedures to Meet Individual Needs

If the suggestions specified earlier are all implemented, the resulting decision aid will be well engineered in terms of an overall user population. However, since the aid's usage is based on the behavior of several individual users rather than a single group, the ability to fine-tune the aid to individual specifications will dramatically improve its acceptance. The more variable individual users (or individual problems) are, the more important this customization will become.

Design-to-Time Control of Processes. Perhaps the most critical variation from problem to problem is the amount of time the user can afford to spend performing an analysis. For high stakes (complex decisions for which time is not a factor), the user would like to ensure maximum validity and completeness, even at the expense of a longer, more extensive analysis; this might entail a variety of sensitivity analyses, consistency checks, data searches, and other procedures. At the opposite extreme, if a decision must be made immediately based only on whatever information is in the decision maker's head, any effort to check for methodological correctness may be perceived as an unnecessary waste of valuable time. Similar variety in users' preferences may stem from the decision makers' personalities, from organizational factors that influence the aid's availability and usage, and from the urgency of other tasks competing for the decision makers' time.

Adaptability to Various Training Levels. One universal problem with multiple-user, interactive computer programs is the need to accommodate a variety of skill and training levels. If an aid is self-explanatory enough to permit error-free use with a complete novice at the controls, it will very likely move far too slowly for a more experienced user. But if the aid is faster and more streamlined (e.g., requiring only abbreviated commands instead of complete words), it is more likely to cause confusion and error in a novice.

Because a decision aid of the sort discussed here should be designed for a wide variety of user skills (it must satisfy a number of naive users, but should also cultivate "repeat customers"), one useful approach might be to provide three tracks:

1. A novice level for the first- or second-time user. This might include a brief tutorial in the aid's procedures, very explicit user instructions with accompanying examples, as natural a mode of interaction as possible, and an analytic capability restricted to a core of basic procedures.

2. A standard level for the occasional user. If the user is experienced enough so that the benefits of the novice level are no longer worth the extra time required, a more streamlined approach might be more effective and might add analytic features beyond the basic novice repertoire.

3. An expert level for the experienced, frequent user. This level would extend the range of analytic capabilities, would emphasize speed and efficiency rather than error protection, and might give the user control of certain performance parameters (e.g., response modes, frequency of reinforcement, output formats, speed/precision trade-offs) to suit individual needs.

It should always be possible for a user to change tracks at any point in the analysis, either permanently or temporarily, without jeopardizing existing results. A help button or instruction could be used to inform the user in more detail about the options available at any point. A more sophisticated aid might keep a record of a decision maker's past usage and performance (e.g., error rates, speed of response), automatically starting the user at the most appropriate level and modifying the level based on current performance (but always subject to user override).

Adaptation to User's Personal Preferences. Once a user has a certain degree of familiarity with the aid, it may be desirable to make minor adjustments and alterations in order to accommodate the user's individual preferences or to comply with a specific set of standard conventions. For example, input-output choices of the color-coding scheme, symbology, and display formatting might initially be given arbitrary default settings, but on the user's request, they might be altered to fit individual needs. Similarly, operational features such as the mode of input or the machine's average time to react may need to be adjusted (as in the case of some computer chess-playing programs, in which a delay was added because users felt uncomfortable with the instantaneous responses the machines had been making).

A more sophisticated approach to customization would have the aid's routines expressed as functions of several parameters, each of which might correspond to some aspect of the user's skills and preferences. The frequent user could initiate a questionnaire routine that would replace the default settings for all of these parameters with user-specified values (e.g., "How good a typist are you?" "Which of these type faces do you prefer?" "In general, which is more important to you, speed or completeness?"). Then, a special version of the aid's routines could be compiled using the profile's values. As those values changed, the user could modify the profile and alter the routines accordingly.

The methods just discussed would require a fairly sophisticated user; a novice or occasional user would not be sensitive enough to minor alterations to make the effort of fine-tuning worthwhile. However, a very sophisticated version of the machine might automatically select which parameters each user could adapt, basing its selection on physiological monitoring, if available, and on the user's behavioral state, inferred from response time, error rates, and answers to direct inquiries (e.g., "Do you want to continue or would you like a break?"). Data about behavioral state could be used to check for user alertness, to regulate the frequency of breaks and reinforcements, and to adjust system parameters experimentally in order to improve user performance.

Conclusion

In the future, methods of computer-aided decision making may bear little resemblance to the methods available today. As the general public becomes more knowledgeable about computers, and as computer usage by nonspecialists becomes widespread, some of the blocks that have been the target of the current efforts may disappear (as others arise). Also, as available technology--both hardware and software--becomes cheaper, more accessible, and more sophisticated, more ambitious goals will become feasible. Speech recognition, natural language comprehension, visual image perception, three-dimensional displays, and even more advanced features will someday be commonplace. However, only by working now to pioneer useful applications can we hope to influence the course of such developments and find a market for them when they are ready. The issues discussed in this section, and the implementation features recommended, will provide a sound basis for decision aid engineering in the near term and a guide for the eventual incorporation of future technology.

INVOLVING USERS IN THE DEVELOPMENT OF DECISION-ANALYTIC AIDS: THE PRINCIPAL FACTOR IN SUCCESSFUL IMPLEMENTATION

Over the past 25 years, hundreds of scientific studies of human judgment and decision making have reached one basic conclusion: Unaided human judgment has limitations.[1] As a result of these findings, as well as of advances in normative decision theory (e.g., von Neumann & Morgenstern, 1947; Savage, 1954; Raiffa, 1968; Keeney & Raiffa, 1976) and computer technology, judgment/decision researchers have begun developing computer-based decision-analytic aids to help decision makers improve and extend their cognitive ability. These aids include different types of multi-attribute utility assessment programs, such as HIVAL (Allardyce & Peterson, 1979) and POLICY (Hammond, Cook, & Adelman, 1977), as well as traditional decision-analytic aids requiring probability and utility assessment, such as INFER (Amey, Feuerwerger & Gulick, 1979a) and OPINT (Amey, Feuerwerger, & Gulick, 1979b). Such aids have been used successfully in a wide range of settings, as indicated in compendiums by Kaplan and Schwartz (1977), by Keeney and Raiffa (1976), and by Kelly (1979).

On the basis of previous success, one can expect increased utilization of computer-based decision-analytic aids with stand-alone capabilities for routine use by internal analysts and decision makers without outside consultation. The thesis of this section is that decision-analytic aids will seldom achieve a stand-alone status unless eventual users are involved in their development. The term <u>users</u> applies here both to the persons running the decision aid and to the decision makers utilizing its results. The previous section focused on the interface between the aid and its hands-on users, who may or may not be decision makers. The position was that the better the general behavioral characteristics of the aid, the higher the motivation of the hands-on user and therefore the greater the probability that the aid

---

[1]The interested reader is referred to Hammond, McClelland, and Mumpower (1980); Slovic, Fischhoff, and Lichtenstein (1977); and Slovic and Lichtenstein (1971) for reviews of this research.

will be successfully integrated into the organization. In this section, the focus is on the interface between the user (and decision aid) and the larger decision-making organization. The position here is that successful general behavioral characteristics of decision aids are necessary, but not sufficient, for aid implementation. In addition, the involvement of decision makers in aid design is essential for implementation, for this involvement develops the understanding and commitment necessary for implementing a different decision-making approach and tailors the characteristics of the aid to the users' needs within their organizational context.

The judgment/decision research literature has not emphasized that user involvement in aid design is important for successful aid implementation. Support for this position comes primarily from two sources: (1) the recent systematic evaluation of an experimental decision-analytic aid developed for use by the Operations Directorate of the Joint Chiefs of Staff and (2) the history of model implementation in operations research and management science. Although the need for user involvement appears obvious, these sources indicated that user involvement is often neglected in the development of analytical decision aids, frequently resulting in unsuccessful implementation efforts.

Evaluation of Decision-Analytic Aids

R-SCREEN. Sage and White (1980) recently evaluated a multi-attribute utility assessment (MAUA) aid, called R-SCREEN (Rapid Screening of Decision Options). The R-SCREEN aid was developed for use by operational analysts in the Joint Operations Division (JOD) within the Operations Directorate of the Office of the Joint Chiefs of Staff.

Funding for R-SCREEN was provided by an agency that is tasked with monitoring, evaluating, and improving the overall information flow within the World Wide Military Command and Control System in support of information reporting, information analysis, decision making, and information dissemination. Specifically, R-SCREEN was developed to support the option generation and selection process as it occurs in command centers in crisis situations.

JOD decision makers used R-SCREEN by implementing four steps. First, they selected from among three prestructured templates (or hierarchies) the template most appropriate for the particular problem at hand, and they made minor modifications to the structure as needed to match the template to the criteria most relevant to the particular problem. Second, they identified various alternative courses of action for evaluation. Third, they scored each of the alternative courses of action on each of the lower level attributes and then assessed criterion importance weights (essentially using Edwards's (1977) ratio estimate technique) in order to determine the relative utility of each alternative. And fourth, they assessed the sensitivity of the analytical results by evaluating the impact of changing utility scores and criterion weights.

R-SCREEN was introduced into the JOD in spring 1979. Decision analysts briefed the JOD staff on how to use the aid, developed a user's guide specifically for the aid (Gulick & Allardyce, 1979), provided on-the-job training sessions throughout the course of the experimental period, and, in general, made themselves immediately available at the request of JOD personnel to discuss R-SCREEN's utilization. JOD personnel were, however, not involved in R-SCREEN's development. The experimental period lasted approximately 6 months.

Sage and White (1980) evaluated R-SCREEN by the following three procedures: (1) informal interviews with JOD personnel and others familiar with the JOD operational environment, (2) study of various written documentation, and (3) detailed analysis of questionnaire responses as well as follow-up interviews with Pentagon personnel and with a group of senior military and civilian students from the Industrial College of the Armed Forces, who were asked to evaluate the aid in an experimental context. Sage and White organized the evaluation responses into the criteria and subcriteria shown in Table 1. Although Sage and White discussed the implications of the responses in terms of each of the 15 subcriteria, they did not give R-SCREEN an explicit score on each criterion. In order to shorten this presentation, Table 1 shows an overall score (+, -, or ?) based on Sage and White's qualitative evaluation on each subcriterion.

R-SCREEN rated extremely well in performance objective achievement (category 1) and efficacy (category 3). These high ratings provide empirical support for the claims of judgment and decision researchers who have argued that decision-analytic aids facilitate clear thinking, educate decision makers about their problems, and facilitate communication (e.g., see Hammond et al., 1980).

R-SCREEN rated poorly, however, in behavioral criteria (category 2). R-SCREEN received a questionable rating on implementability because of participants' reservations concerning its usefulness in a crisis management environment. These reservations relate directly to R-SCREEN's ratings on political acceptability and institutional constraints. Sage and White discuss these two subcriteria as follows:

> Political acceptability:
>
> Political issues were viewed by several subjects as potential barriers to acceptance of systemic aids, such as R-SCREEN, into an operational environment. Lack of senior level receptivity and the personal decision-making styles of flag officers were seen as potential hindrances. Full management and other leadership commitment to implementation testing of decision aids were viewed as very necessary. Significant barriers to acceptance of an aid were felt to result with the absence of this commitment.
>
> Institutional constraints:
>
> Questionnaire responses indicated a concern that R-SCREEN does not directly address the needs of the JOD, is not particularly well matched to the behavioral characteristics of the operational environment, may not enhance information flow, and does not possess desirable time to use response characteristics for typical JOD operations. (1980, p. 0.11)

Table 1

Criteria for Decision Aid Evaluation and Evaluation Scores for R-SCREEN

Criterion	Rating
Algorithmic effectiveness or performance objective achievement	
Logical soundness	+
Improved decision quality	+
Decision process changes	+
Behavioral or human factors	
Political acceptability	-
Institutional constraints	-
Implementability	?
Procedural changes	-
Side effects	?
Efficacy	
Time requirements	-
Leadership and training requirements	?
Communication accomplishments	?
Educational accomplishments	+
Documentation	+
Reliability	+
Convenience of access	+

Note: + means performed well; - means performed poorly; ? means performed well and poorly on questions comprising the subcriterion category.

In short, R-SCREEN was not tailored to the personal needs and organizational context of its eventual users. As a result, evaluation responses indicated that its implementation in JOD was questionable--and, in fact, it has not been implemented to date even though respondents believed it would improve decision quality, just as its designers had claimed.

It is important to contrast the above unsuccessful implementation effort with a successful one in order to gain insight into the extent to which implementation is enhanced by tailoring a decision-analytic aid to the personal needs and organizational context of its eventual users. The authors, however, are not aware of any evaluation of a decision-analytic aid in its operational context that is as systematic and thorough as the evaluation conducted by Sage and White. Although post hoc evaluations of successful implementation efforts are open to charges of bias, such an evaluation is presented briefly in an effort to help readers evaluate the adequacy of the thesis advanced in this section.

MCCRESSA. The U.S. Marine Corps, as well as other services, has a continuing problem in assessing, under peacetime conditions, the combat readiness of combat units. The problem is compounded by the many heterogeneous attributes that are used to describe the performance of individual combat units and by the many criteria, both objective and subjective, that are commonly used by force commanders to define a successful level of combat readiness. Historically, there has been almost no acceptable standardization or formalization of the process of combat readiness evaluation or validation of evaluation results.

In support of the Marine Corps Combat Readiness Evaluation System (MCCRES), an MAUA aid called Marine Corps Combat Readiness Evaluation System Software Application (MCCRESSA) was developed (Allen & Allardyce, 1976). The Marine Corps successfully tested MCCRESSA in an operational setting in August 1977 and the aid is now in routine use throughout the Marine Corps.

MCCRESSA and R-SCREEN are extremely similar. Both were designed from the same generic MAUA software. If anything, R-SCREEN is more sophisticated analytically than MCCRESSA because it forces the user to assign criterion weights moving from the bottom to the top of the hierarchy, thereby ensuring that the upper level weights are determined by the scores on the lower level attributes and not by the user's general perception of the relative importance of the upper level attributes. Both aids were designed to have stand-alone capabilities. Yet, MCCRESSA was successfully implemented and R-SCREEN was not.

MCCRESSA was successfully implemented because its eventual users were involved throughout the entire process of development and implementation. The decision aid analysts worked directly with the five Marine colonels tasked with developing and implementing MCCRES over a 1-year period. These men decided on the criteria, hierarchical structure, and weights in the MAUA model within MCCRESSA. They decided how inputs to MCCRESSA would be made during actual MCCRES evaluations. They decided on the type of outputs MCCRESSA had to provide, and the constraints under which these outputs would have to be provided, within their operational context.

After a prototype aid was developed, the Marine colonels chaired a 2-day conference for all field commanders who would participate in MCCRES evaluations. They showed the commanders how MCCRESSA would be used during each evaluation and gave them an opportunity to ask questions, raise concerns, and suggest ways of better tailoring MCCRESSA to the evaluation process. The colonels also went to each of the Marine bases, where evaluations were held to answer questions and obtain suggestions from personnel who would actually use MCCRESSA during an evaluation. Some of the lower level attributes in the MAUA hierarchy and some of the procedures for using MCCRESSA were modified on the basis of the concerns and suggestions raised during the conference and tour. There were additional minor modifications of MCCRESSA after its initial application during some MCCRESS evaluations.

In sum, user involvement throughout development and implementation ensured that MCCRESSA was tailored to the Marine Corps's needs and organizational context. The authors believe this to be the principal factor in MCCRESSA's successful implementation in the Marine Corps.

Implementing Operations Research Models

Although there are distinct differences between decision-analytic aids and operations research models, both represent highly analytic techniques for assisting the process and quality of decision making within large organizations. Consequently, both face similar implementation problems. The operations research (OR)/management science (MS) literature over the past two decades has (1) documented numerous cases in which clients have not used analytically rigorous OR models developed for them, (2) tried to explain this phenomenon, and (3) offered suggestions for minimizing unsuccessful implementations. This section briefly reviews this literature.

Ginzberg (1978) divided the OR/MS literature on implementation into two types: the normative approach (e.g., Ackoff, 1960; Argyris, 1971; Grayson, 1973) and the factor approach (e.g., Drake, 1971; Powers & Dickson, 1973; Rubenstein, Radner, Baker, Heiman, & McColly, 1967).

> The normative approach is based on the field experience of a number of MS researcher/practitioners. These researchers typically looked back at one or more cases they were involved in where there was substantial implementation difficulty, and attempted to draw from these experiences the general nature of implementation problems and their solutions. Looking at this literature in aggregate, we find substantial disagreement on just what the solution to implementation problems should be.... The next development in implementation research was the factor approach. Each factor study begins by identifying a group of variables potentially relevant to implementation outcomes. Data are then collected from a sample of MS implementation projects - some successful and others not - and are used to assess the relative importance of the different variables (or factors) to implementation outcomes. The results, however, are rather disappointing. Few general guidelines have emerged from this research, the results of different studies being contradictory in a number of cases. The only result which is firmly established by this research is the importance of management support and user involvement to the successful implementation of MS/MIS projects. [emphasis added] (Ginzberg, 1978, pp. 57-58)

The research by Lonnstedt (1975) and Shycon (1977), which was not cited by Ginzberg (1978), further supports Ginzberg's conclusion. Lonnstedt (1975) interviewed key operations personnel in 12 companies, each with its own OR division, listed in the Stockholm Stock Exchange in an effort to identify factors related to the implementation of operations research solutions. The study sample was composed of 107 OR projects proposed for implementation; 29 of the projects were not implemented by the user.

The results of the survey are presented in Table 2. There is a positive relationship between implementation and (1) the user's collaboration in defining the problem, (2) problem characteristics, and (3) the value the user places on the proposed OR solution. All three factors require continual interaction between the user and the OR modeler throughout the process of model development.

Table 2

Summary of Survey Results: Relationships Between Variables and Proposal Implementation (from Lonnstedt, 1975)

Variable group	Variable	Chi square	Significance (p)	Influence of nonresponses on conclusion
Collaboration	User's participation	19.1	$< .001$	May influence
	Initiator of project	16.7	$< .001$	May influence
Characteristic of problem	Problem limitation	52.5	$< .0001$	No influence
	Quantifiability of variables	24.4	$< .0001$	No influence
	Availability of data	33.0	$< .0001$	No influence
Proposal value and cost	Value of resultant solution	34.1	$< .0001$	May influence
	Internal charging	0.85	$< .4$	May influence

Shycon (1977) conducted two surveys of large OR projects varying in their degree of successful implementation. Both surveys categorized each OR project team's organization as one of the following three types: (1) the wholly management science team consisting entirely of management science personnel, with minimal interaction with others; (2) the management science team with marginal communication to management, largely at the middle management level, through frequent reporting; and (3) the interparticipative management science management team, which involves a working partnership of members of the management science group and middle and upper management representing both line and staff functions. Under this classification system, the project team for implementing R-SCREEN was a wholly management science team. In contrast, the team for MCCRESSA was an interparticipative management science/management team.

Table 3 presents the results of Shycon's (1977) surveys. The wholly management science team achieved the lowest degree of implementation success. The MS team with communication and the interparticipative team consistently achieved high levels of implementation success. Neither of the latter two organization types, however, showed any distinct advantage over the other. Given the greater cost of the interparticipative team, the results suggest that the MS team with communication is the most cost-effective project team organization. Nevertheless, the results support post hoc the actual outcomes for R-SCREEN and MCCRESSA regarding implementation success.

Table 3

Project Team Organization and Project Success (from Shycon, 1977)

Type of project	Type of firm	Approx. project cost[a]	Degree of success[b]
	1971		
1. Wholly management science team			
Determination of regional distribution requirements	Regional food distribution	36	Partial
Sales forecasting and inventory planning system	National tool manufacturer	68	Partial, long-term
Sales forecasting and inventory planning system	Division of major drug manufacturer	36	Little immediate, some longer-term
Determination of service call response strategy and facilities required	Regional public utility	30	Complete
Determination of service call response strategy and facilities required	Regional public utility (different from above)	32	Partial
2. MS team with marginal communication			
Design of national distribution system	National food processor	140	Complete
Design of total logistics system, manufacturing plants, and distribution	National food processor	95	Complete
Service facility requirements	Heavy machinery manufacturer	55	Complete
Integrated distribution requirements for diverse divisions	Major drug and toiletries manufacturer	54	Minor benefits
3. Fully interparticipative MS team			
Total management planning program: procurement, inventory, scheduling, distribution sales forecasting, and marketing planning	Major national meat packer	[illegible]	Complete
Design of national distribution system	National food processor	12[illegible]	Complete
Design of national distribution system	Major instrument and supplier manufacturer	96	Complete
Basic simulation of company operations for management policy testing	Major pharmaceutical manufacturer	104	Complete
Evaluation system for R&D projects and tool program for rank order and funding	Major synthetic fiber	160	Partial
Corporate strategy model for decisions in marketing, manufacturing, and capital investment	Major integrated paper products manufacturer	250	Complete
Design of total logistics system, manufacturing plants, and distribution	National industrial products plastics manufacturer	75	Complete
	1977		
1. Wholly management science team			
Determination of national distribution and transportation strategy	Consumer large instruments	42	Little
Multiplant manufacturing operational strategy	Building products manufacturer	77	Little
2. MS team with marginal communication			
Research and development of inventory planning system	Sporting goods manufacturer	11	Complete
Development of marketing strategy with attendant distribution requirements	Automobile after market manufacturer	65	Complete
Queuing simulation model, flow shop	Principal furniture manufacturer	55	Partial
Evaluation of marketing channels and design of order entry	Food manufacturer	80	Complete
3. Fully interparticipative MS team			
Simulation and design of national distribution system	Chemical processor	98	Partial
Evaluation of customer service requirements and design of national distribution system	National food processor	132	Complete
Design of integrated strategy planning model	International extractive and and fabrication company	108	Partial
Development of marketing channels and supply strategy	High technology industrial product company	94	Complete
Development of specifications for inventory management system and implementation	Principal food processor	265	Complete
Design of national distribution system	Industrial products hard good manufacturer	85	Complete

[a] Includes all services, internal personnel, and external expenses incurred. All figures for 1971 adjusted to 1971 dollars. All figures for 1977 adjusted to 1977 dollars.

[b] Degree of success is necessarily partly subjective; however, measurable criteria are as follows: little--less than 25% implemented, benefits did not justify cost of study; partial--25-60% implemented, identifiable benefits yield return on project investment (ROPI) less than 100% per year; complete--greater than 60% implementation, ROPI greater than 100% per year.

Why User Involvement Is Essential

User involvement throughout the development of decision-analytic aids is essential to make users comfortable with a decision-analytic approach to decision making. User involvement is also essential if the analysts are to learn enough about the users' goals and working environment to tailor the decision-analytic aid to the users' personal needs and organizational context.

Decision makers are not decision analysts. Although they may identify important factors for the decision at hand, they will seldom build a decision tree or multi-attribute hierarchy. Nor will they typically quantify the probability of uncertain events or the relative importance of attributes over their range of variation for the set of alternatives, or calculate expected utilities to determine the preferred action. Since formal decision analysis is not their standard mode of decision making, decision makers need to learn basic decision-analytic concepts and feel comfortable providing the aid's required inputs and interpreting its outputs before they will use it routinely. Interaction between the analyst and the decision maker is essential to this learning process and, more generally, to developing the confidence and commitment necessary for implementing a different approach to decision making.

The interaction must be a two-way process, however, for the analysts must understand the organization's broader goals, working environment, and available resources in order to develop an effective decision-analytic aid with stand-alone capabilities. For example, the analysts must understand how the decision makers want to use the aid in order to design its output so that it most effectively meets the decision makers' goals. The analysts need to understand the different tasks required to achieve these goals, the different types of people who will perform these tasks, and the factors that facilitate or limit task accomplishment in order to design the aid so that it not only fits into, but improves, the working environment. And the analysts need to know the expected fiscal resources available for operationally utilizing the aid so that it is designed cost-effectively. Involving the user throughout the development process increases the probability that analysts will obtain such information about the organization.

Knowledge about the working environment within which an aid will be used permits analysts to tailor the aid to the users' personal needs in their organizational context. The need for such tailoring has become increasingly documented in the OR/MS literature. For example, in an analysis of implementation of risk analysis methods, Carter (1972) found that unsuccessful efforts tended to have analyses performed by central staffs responsible to corporate rather than divisional managers. Division managers perceived a divided loyalty within the management science staff; this perception resulted in a breakdown in trust and cooperation. Wolek (1975) cites a case in which a rational system for selecting R&D projects was formally adopted but not used because the technique conflicted with the highly personal leadership style of the company's president. And the R-SCREEN example illustrates that a decision-analytic aid that is beneficial will not necessarily be adopted; the aid also has to score well on behavioral characteristics.

A decision-analytic aid will change the users' working environment, for it will alter the organization's decision-making process. This change will occur whether, according to Von Winterfeldt's (1979) classification, the aid is (1) a highly specific aid with a previously determined structure and stored data (e.g., weights) like MCCRESSA, (2) a multipurpose aid with no substantive structure and stored data, or (3) an aid like R-SCREEN that combines features of both extremes. However, the extent of the change and its effect on the interpersonal relations and functions of different people within the organization will vary. In addition, it may be necessary to hire (or train internally) skilled decision analysts to use effectively multipurpose aids with no previously determined structure. Such aids require more operational support than either of the other two types because they are not used for repetitive decisions. In contrast, structured aids used for repetitive decision making, like MCCRESSA, require little operational support because their analytical structure is designed primarily to implement an existing on-going process more effectively. Nevertheless, the smoother the expected and actual transition to a new decision-making approach and working environment, the greater the probability of successful aid implementation. User involvement throughout the process of aid development and subsequent implementation, that is, "from initial planning and feasibility testing through installation and evaluation" (Ginzberg, 1978, p. 59), can ease this transition greatly.

In closing, it is interesting to note that the need for user involvement in aid development may not surprise many decision analysts and researchers. What may surprise them is the empirical support for this need; there are many cases in the OR/MS literature, and now some in the decision-analytic literature, in which users have not been involved in aid development. Two possible reasons for this state of affairs come rapidly to mind. First, many decision analysts and operations researchers have not realized the importance of user involvement for successful implementation. It is one thing to give the concept of user involvement lip-service and quite another to consider it the principal factor in successful implementation. Second, ensuring user involvement is a difficult task. Users often fail to appreciate the importance of their involvement and, as a result, consider aid development to be solely the job of the analyst instead of a two-way interaction. The primary usefulness of this section may well lie in alerting analysts to the necessity of making users realize the importance of their involvement in aid development.

DECISION AID EVALUATION

The preceding sections have been concerned with the problem of designing effective decision aids, but the present section turns to the problem of evaluating decision aids. Although design and evaluation are highly interrelated, they are approached from different perspectives and therefore involve somewhat different difficulties. Design often begins in the absence of an implemented system and must determine how to incorporate a variety of capabilities. Evaluation begins with a partially or completely implemented system and must determine whether the system does what it is meant to do. Optimally, the two activities are performed iteratively, with the answers to design questions posing evaluation questions, and the answers to evaluation questions posing design problems. Nonetheless, they are quite distinct activities.

Three major types of questions relate to the evaluation of decision aids; these questions correspond to the three types of interfaces identified in the introduction to this report (Figure 1). First, an evaluation may attempt to answer questions about the aid's compatibility with its immediate users. Such questions are concerned with the human factors of the aid; for example: Are its displays effective? Is it tedious to use? In addition, the decision aid/user interface is the point at which questions about the comprehensibility of an aid's underlying model are addressed.

A second type of evaluation occurs at the interface between the user (and the decision aid) and the remainder of the decision-making organization. This interface poses questions about the collectibility of the aid's required inputs and the communicability of its outputs. An aid that is comprehensible only to its immediate user is likely to be useless. The decision-making approach used by the aid must be integrated into the larger decision-making organization.

Finally, a third type of evaluation is appropriate for the interface between the decision-making organization and its environment. At this point the ultimate question of the aid's effectiveness comes into play; namely, has the aid improved the organization's output or performance. Similarly, there are questions about the range of environments or problem areas over which the aid provides improved organization performance.

These three types of interfaces--decision aid/user, user/organization, organization/environment--are by no means independent. In fact, they are nested: User/organization effectiveness is necessarily influenced by aid/user effectiveness, and organization/environment effectiveness is necessarily influenced by the effectiveness of the other two interfaces. Nevertheless, the three types of interfaces do have different implications for evaluation, which justifies their use as a framework for discussing aid evaluation.

The subsections that follow use this framework to examine several aspects of evaluation. The first subsection considers the problem of identifying measures of effectiveness. The second addresses the selection of a setting for conducting the evaluation. The third discusses the selection of a method for data collection. And the fourth discusses the question of what is being compared in the evaluation.

Measures of Effectiveness

If an evaluation is to be effective, the evaluator must decide in advance what is to be examined. This is done by identifying one or more measures of effectiveness that are designed to answer the evaluator's questions. Ideally, these measures of effectiveness are objectively measurable and quantitative variables that will describe the effectiveness of the aid. In the present case, however, the term measure of effectiveness will also include subjectively measurable variables and variables that result in qualitative rather than quantitative descriptions. The only restrictions are that the variable must be measurable and that it should be expected to correlate (positively or negatively) with the effectiveness or efficiency of the aid.

Although it would be impossible to list the measures that are appropriate to all evaluations, another approach may provide some insight. The process by which an aid is used can be viewed as a series of stages that proceed from initial data collection to decision implementation (see Figure 2). Under this assumption, anything that improves the effectiveness or efficiency of one stage should improve the overall effectiveness. Thus, the discussion of measures of effectiveness can be simplified somewhat by discussing how the aid can affect each stage.

Data Collection. Data collection is the stage during which the decision-making organization extracts information from its environment. In a military context, this is the domain of intelligence. In the government, it is the domain of, for example, the Census Bureau or the Bureau of Labor Statistics. In the present context, it is part of the organization/environment interface.

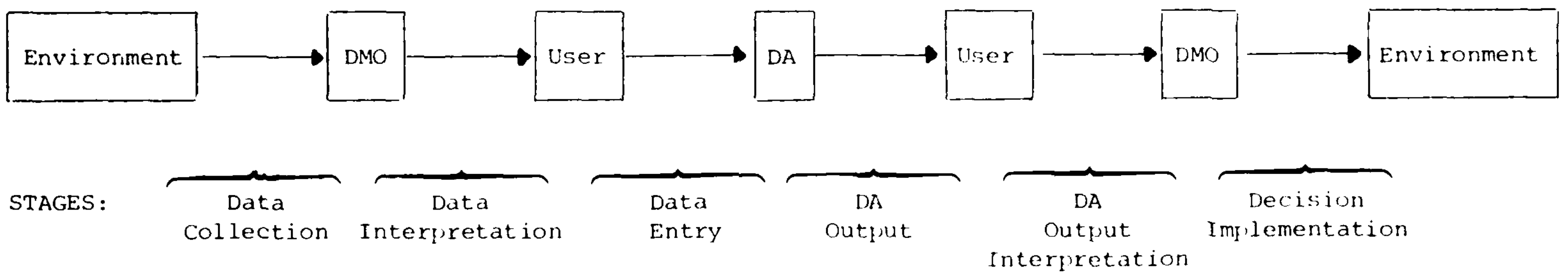
At first glance, this stage might seem remote from the aid and therefore not subject to its influence. Nothing could be further from the truth. The introduction of an aid has the potential to greatly improve data collection by demanding a better organized and better quantified measurement of the environment. In addition, the aid could possibly directly interface with the sensing equipment, which could, in some instances, eliminate errors that might otherwise be introduced as the data are transferred through the organization to the user and into the aid.

A decision aid could, however, have a negative impact on the data collection stage. If the model underlying the aid were either inaccurate or unintuitive, it could compel the collection efforts to be misdirected. Also, an increased data collection effort might be required, thereby increasing costs.

Data Interpretation. The data interpretation stage is the stage during which the members of the organization transform and otherwise interpret the information about the environment. This process may follow strict procedures or it may involve subjective judgments concerning the implications of the data. Although this process may take place throughout the organization, it is considered here as part of the user/organization interface, since this is the point at which the final judgments must be made.

An aid could improve the data interpretation stage by improving the organization's ability to focus on critical information. By disaggregating a problem into meaningful and manageable subproblems, the aid may indicate how the data should be organized, how the data should be transformed, or what types of judgments will be required. Also, the aid may compel a more careful identification of options than is normally undertaken.

In contrast to these benefits, an aid may introduce substantial costs to the data interpretation stage. The aid might require specially trained personnel; it might increase the overall workload by demanding inputs that would not otherwise be collected; and it might create a strain by requiring



DA = decision aid
DMO = decision-making organization

Figure 2. Stage model of decision aid usage.

organization members to think in a manner that is neither natural nor intuitive.

Data Entry. The data entry stage is the period during which the user provides the aid with its required inputs. This stage is part of the aid/user interface.

Data entry is a necessary and frequently arduous aspect of decision aid usage. For this stage an aid will be evaluated on the ease with which the data entry can be performed. The aid should make data entry as rapid as possible and permit a wide range of editing options. The aid should enable untrained personnel to perform data entry, and it should provide easily understood prompts to help the user accomplish this task. Finally, the decision aid should place minimal psychological discomfort on the user. These points are discussed in detail earlier in this report.

Decision Aid Output. The output stage is the period during which an aid provides a user with its results. This stage is clearly part of the aid/user interface. The user for this stage, however, need not be the same individual as the user in the data entry stage.

During the data output stage the aid has its most obvious and immediate opportunity to be of value to the decision-making organization. To accomplish this the aid must provide rapid, thorough, and effective interaction with the user. Sensitivity analyses are critical, because they inform the user of the aspects of the decision that require further inspection. Moreover, the displays must be both accurate and interpretable so that the user will readily understand the underlying model and why it has provided the displayed results.

These benefits of an aid will not be without cost, including the cost of the equipment itself. Usage may, at this stage more than any other, require a specially trained user. This is the point at which a careful understanding of the underlying model and the options available for its exploration can pay off. The job of conveying the model's results to the organization will fall upon this user.

Decision Aid Output Interpretation. The output interpretation stage is the period during which the implications of the aid's analysis are conveyed to the organization and a decision is made. Although this stage involves many actors in the organization and may only briefly involve the user, it is considered, nonetheless, to be part of the user/organization interface.

The decision aid's influence on this stage will depend on the aid's ability to structure and organize the problem to which it was applied. If the model is conceptually complete, coherent, and rational from the point of view of the decision makers within the organization, then its results have a chance of acceptance. Especially important are the communicability and justifiability of the results' implications. If, however, anyone in the chain of communication leading to the decision makers or the decision makers themselves feel annoyed at or uncertain about the results, then these results are likely to be ignored and possibly suppressed from that point on. One way to minimize this problem is to have users (both hands-on users and decision makers) involved in designing the decision aid. This develops the understanding and commitment necessary for implementing a different decision-making approach, and the characteristics of the decision aid can be tailored to the needs of users within their organization.

Decision Implementation. The decision implementation stage is the period during which a decision is translated into some action on the part of the organization. In the case of operational decisions, this stage is part of the organization/environment interface. Internal development decisions, those directly affecting the organization, are not actually excluded from the framework, but there is no interface required with the environment. Instead, the organization should be depicted as feeding back on itself in a self-regulatory fashion.

For this stage the most fundamental question is whether the aid has led to a sound decision. If the decision's implications are correct and they are not ignored, then the aid has provided its major benefit. The aid can, in addition, provide insight into how the decision should be implemented and what is likely to occur following implementation. Even at this stage, however, a correct decision could be undermined, if it is both counterintuitive and unjustifiable. Thus, the communicability and comprehensibility of the aid must carry through even to this late stage of the process.

Summary of Potential Measures of Effectiveness. To provide a summary of some potential measures, Figure 3 reiterates many of the issues raised in the preceding discussion. In this figure, the issues are organized into a hierarchy, and it is assumed that each terminal node could be translated into a measure. This representation is not meant to advocate any particular set of measures or any particular approach to evaluation. It is simply a summary of a number of measures that may be relevant to any specific decision aid evaluation.

Organizational Impact. Although it is implicit in the comments of the preceding discussion, one additional point deserves mention. The introduction of an aid into a complex organization is unlikely to be accomplished without changes in the organization. A decision aid is not like a new stereo component that can simply plug into the old system as a replacement for some older component. For one thing, the aid will be unlike any existing component, that is, individual, of the organization. It cannot truly replace a person, because it cannot do all that a person does. This fact implies that the introduction of the aid will necessarily redefine certain roles within the organization.

The impact of a decision aid on an organization may in fact be sweeping. To use an aid, new channels of communication may be required and new areas of authority may need to be defined. Such changes could be minimized by a careful design effort prior to development of the aid. Nonetheless, some organizational change is likely to be necessary. In the event that the change is too great, one can expect the aid to lie idle. However, if the organizational changes are slight or at least carefully planned, the aid will have a chance to take hold and contribute to the effectiveness of the decision-making organization.

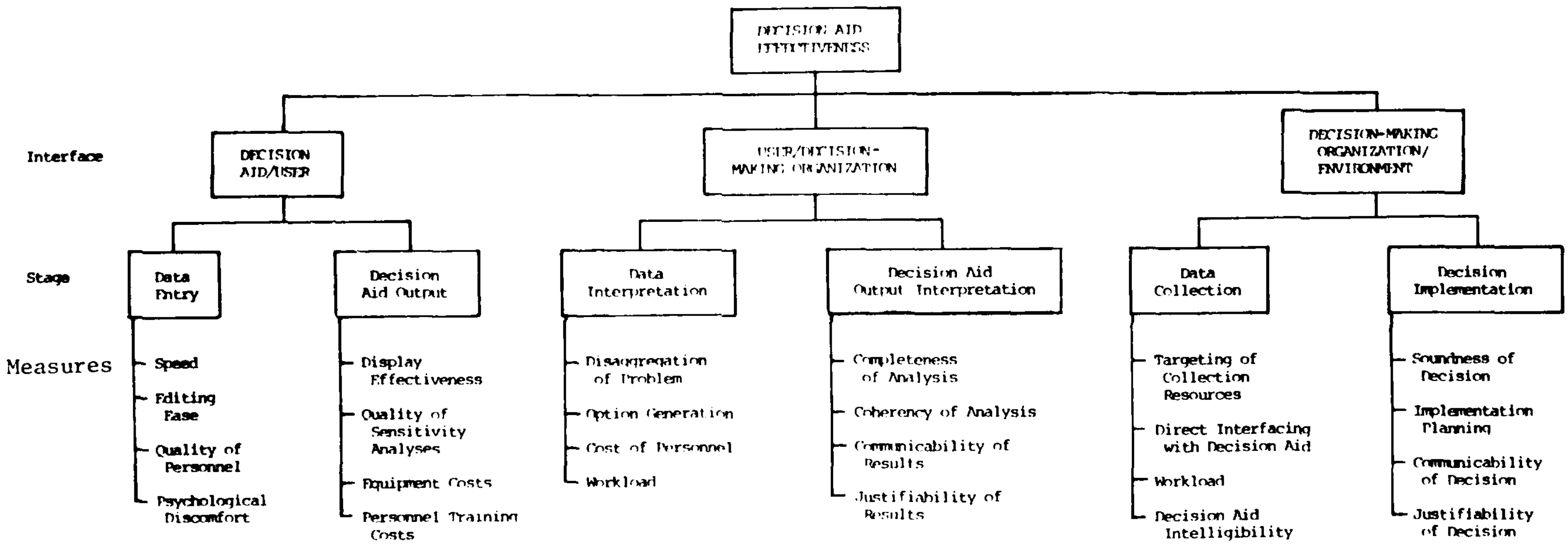


Figure 3. Summary of potential measures of effectiveness.

Settings for Decision Aid Evaluations

Before the measures of effectiveness for a decision aid can be collected and analyzed, it is necessary to construct a setting in which the aid can be operated. The setting might simply be a laboratory experiment with a mock problem, or it might be a full field trial. Such settings differ in terms of their fidelity, or similarity, to the expected operational setting, the amount of experimental control that they provide, and their costs. Thus, the choice of a setting can be a difficult one.

Figure 4 depicts the situation that prevails when one attempts to conduct an assessment of effectiveness. The first part of the figure, labeled Target Setting, represents the expected operational setting for the aid. Of course, this setting will not be available for evaluation purposes unless the aid is actually deployed. In lieu of the target setting, it is therefore necessary to construct a test setting within which the evaluation can proceed.

One of the most fundamental dimensions over which test settings can vary is their degree of fidelity to the target setting. The simulated environment, the simulated organization, and even the simulated user can range between being only superficially accurate to being accurate in great detail. By itself, high fidelity is, of course, desirable in any evaluation setting, but it is expensive. In addition to increased dollar costs and evaluation time, fidelity introduces a cost in terms of loss of experimenter control. This means, on the one hand, that it may be increasingly difficult to obtain the desired measures and, on the other hand, that these measures will be increasingly susceptible to influences that are extraneous to the evaluation context. Even if one is successful in eliminating extraneous influences from the evaluation, there will be increased difficulty in specifying and controlling causal relationships in a high-fidelity setting. Thus, a trade-off is established between fidelity and costs such that it is desirable to simulate only as much of the target setting as is necessary to support a particular inference.

This concept of fidelity and the trade-offs it implies can be further examined by considering four settings defined by whether they involve high or low fidelity for the organization and high or low fidelity for the environment. The additional settings provided by low fidelity for the user are not examined because a qualified user is required for any evaluation.

Of the four settings, the setting with a low-fidelity environment and low-fidelity organization is the most austere. Such a setting is well suited to decision aid design questions concerning user compatibility. These questions are primarily concerned with the aid/user interface and therefore need not concern the organization or even the true environment. Since so little simulation effort is required, this setting can be implemented in a carefully controlled laboratory experiment.

The setting with a high-fidelity environment and low-fidelity organization can also be implemented in the laboratory, but it serves a different purpose. This setting focuses on providing the user with realistic data about the environment, realistic options, and realistic scenarios, but the organization through which the user would interact with this environment is only superficially implemented. Thus, it is possible to investigate the

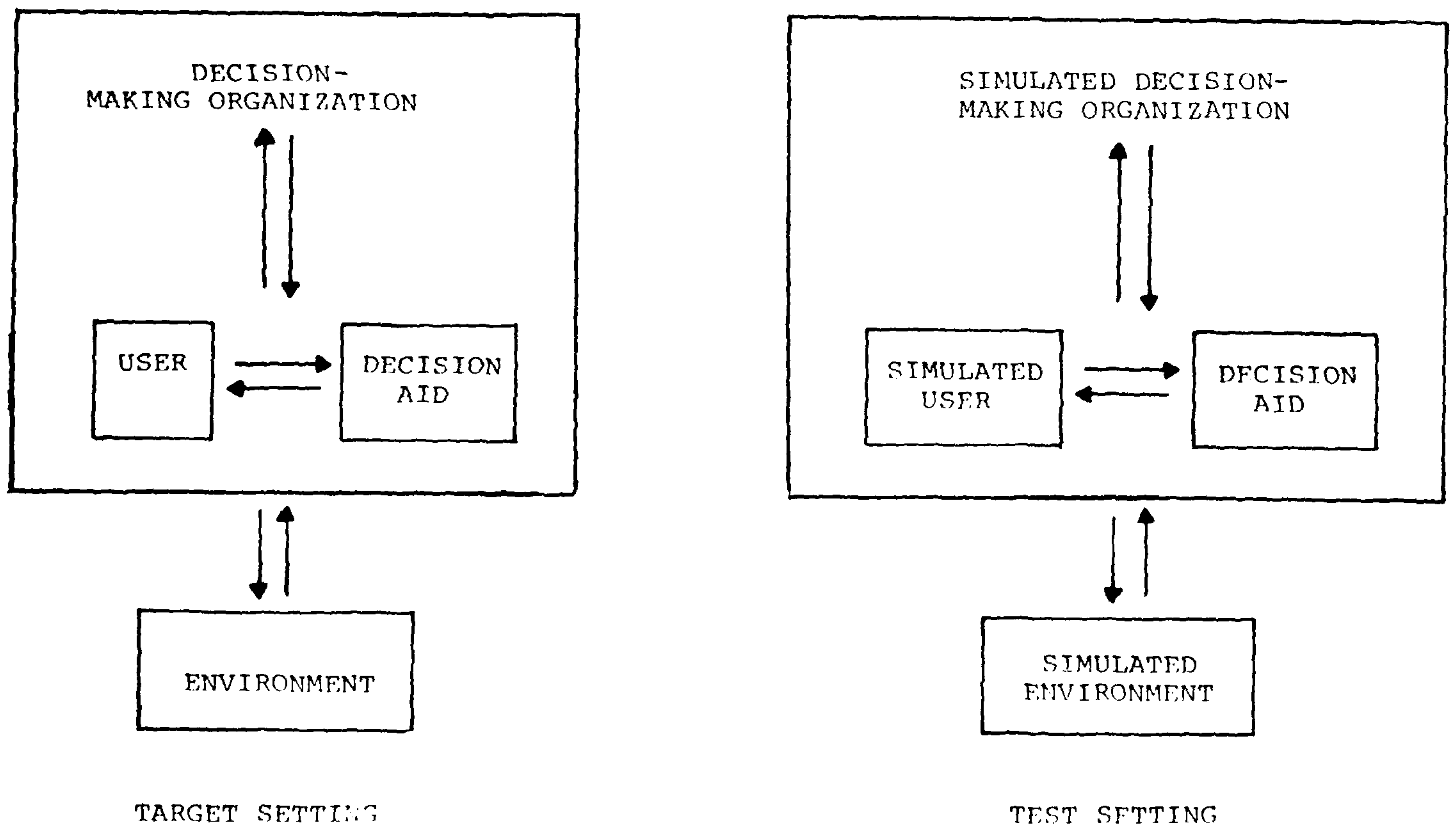


Figure 4. Notional representation of the setting for a decision aid evaluation.

coherence and completeness of the model underlying the aid without going the additional step and ascertaining whether the aid will improve organization performance. Such an analysis really evaluates only questions related to the aid/user interface.

The setting with a low-fidelity environment and high-fidelity organization provides the means to answer questions about the user/organization interface. In this setting, little concern is devoted to constructing realistic problems, and a great deal of concern is devoted to simulating the lines of communication and authority within the organization. Since a simulated organization is outside the scope of most laboratories, this type of setting is better thought of as a gaming simulation. Although the departure from the laboratory is necessary to evaluate the user/organization interface, this setting implies decreasing control and increasing costs.

Finally, the setting with a high-fidelity environment and high-fidelity organization is the most accurate, virtually requiring a field test with a realistic and well-implemented problem scenario. This accuracy is obtained at a high cost, but it is necessary to fully answer questions about decision aid effects on organization performance. Since these questions about the organization/environment interface are the ultimate questions concerning aid effectiveness, field tests of this sort are a highly desirable precursor to aid deployment.

Table 4 summarizes the comments of the preceding paragraphs. Clearly, the choice of an evaluation setting interacts with the type of question that one hopes to answer. In light of these fidelity/cost trade-offs, the following approach to evaluation seems justified.

If all types of questions are important, investigate them in the following order: questions of user compatibility, questions of aid coherence and completeness, questions of aid compatibility with the organization, and questions of aid effect on organization performance. Although the order of the first two evaluations may change, cost considerations are likely to compel the remainder of this evaluation strategy, since it will be desirable to have suffered the least costs in the event that any one evaluation provides a negative result.

Methods for Collecting Measures of Effectiveness

There are three major methods for obtaining measures of effectiveness: objective measurement, subjective judgment, and expert observation. The first of these is the most familiar and is most associated with experimentation and the scientific method. The second technique, subjective judgment, involves requiring users or other participants in the experiment to score their experiences, usually via a questionnaire following the experiment. The final technique, expert observation, also involves subjective judgments, only on the part of nonparticipating observers of the experiment. Although there is a prevailing prejudice in favor of objective measurement, all of these methods can be valid provided they are properly employed.

In decision aid evaluation, objective measurements are likely to consist of speed and frequency measures. It will be important to know how long

Table 4

Summary of Alternative Evaluation Settings

Setting fidelity	Type of evaluation	Interface examined	Question evaluated	Cost	Experimenter control
Low environment Low DMO	Laboratory	DA/U	DA user compatibility	Low	High
High environment Low DMO	Laboratory	DA/U	DA coherence and completeness	Low	High
Low environment High DMO	Gaming simulation	U/DMO	DA compatibility with DMO	Moderate	Moderate
High environment High DMO	Field test	DMO/ENV	DA effect on DMO performance	High	Low

Note: DMO means decision-making organization; DA means decision aid; U means user; and ENV means environment.

some process requires or how frequently errors occur. Table 5 suggests ways in which such objective measurements could be used to evaluate the aid/user, user/organization, and organization/environment interfaces.

No less important than objective measures are the assessments of a participant's satisfaction, complaints, or other judgments about an aid's effectiveness. Not only are these judgments easier to collect than objective measures, but they represent a class of data that is a critical determinant of an aid's ultimate acceptability. An aid that is objectively effective, yet subjectively unacceptable, is still unacceptable, since its chances for effective deployment are low. Thus, subjective judgments should not be ignored. Table 5 suggests several measures of effectiveness that measure subjective judgments at each of the three interfaces.

The final method of evaluation, expert observation, differs from subjective judgment in that the raters are not participants in the experiment but outside observers. Judgments of this sort can be critical for answering questions about the completeness or soundness of a decision, because the correct decision is unlikely to be known. In the absence of any objective definition of correctness or accuracy, expert judgments must suffice. In this capacity the experts play the same role that a coach or trainer plays. Experts are deemed correct by virtue of their greater experience. Table 5 suggests some w. s that expert observations can assist an evaluation effort.

I summary, each of the three methods of collecting measures of effectiveness plays an important role at all three interfaces. Objective measurements can provide an understanding of the frequency and speed with which particular events occur. Subjective judgments can provide information about the decision-making process from the participants' perspective. And expert observations can provide a notion of decision soundness and accuracy, when such objective definitions are usually unavailable.

What Is Being Compared?

Before any evaluation can proceed, it is necessary to ask what is being compared. In a formal experiment, the comparison is between a test and a control condition. Similarly, some notion of a control condition or at least a contrasting condition is required for aid evaluation.

Consider the three interfaces once again. Evaluations involving the aid/user interface are largely concerned with user compatibility and aid coherence. Evaluations could reasonably compare alternate configurations of the aid or even different aids. A comparison between an aid and no aid is, however, inappropriate. Questions about the aid/user interface assume a decision aid, just as they assume a user.

Evaluations involving the user/organization or organization/environment interfaces permit more comparisons than does the previous type of evaluation. In particular, at these interfaces it is poss''le to examine the organization's operation both in the presence and in the absence of the aid. To perform such an evaluation one must recognize, however, that the definition of the user--and, therefore, the interface--will change when the aid is not present. In other words, an aid is not like a plug-in module; its introduction

Table 5

Measures of Effectiveness for Each Method at Each Interface

Method	Measure of effectiveness		
	Decision aid/user interface	User/decision-making organization interface	Decision-making organization/environment interface
Objective measurement	Time for data entry	Time for data interpretation	Speed and accuracy of decision implementation
Subjective judgment	Rating of aid display effectiveness	Rating of communicability of aid outputs	Rating of comprehensibility of decision
Expert observation	Rating of completeness of aid analysis	Rating of soundness of data interpretation	Rating of soundness of decision

will necessarily require changes in the organization. Thus, an effort to compare performance when an aid is present with performance when the aid is absent may be confounded by effects of the organizational changes required by the aid.

While effective performance is likely to be welcome regardless of whether it is caused by the aid or by the changes that the aid requires, this confounding should not be overlooked. The primary concern is that the bulk of the benefit attributed to an aid not be due to its concomitant organizational changes. If it is possible that the major benefit is due to organizational change, thorough examination requires that the evaluator attempt to compare performance in the modified organization in the presence and in the absence of the aid.

Summary

Decision aid evaluation begins with a recognition of the fact that the aid will simply be one component of a more complex information-processing system. The inputs to the aid and the outputs from the aid will probably travel through many layers of a complex decision-making organization. Thus, a thorough analysis of decision aid effects is likely to examine aspects of organization performance as well as aid performance.

In examining complex information-processing systems of this sort, it is useful to concentrate on the interfaces between the system components. These are the points at which the system reveals itself. Bottlenecks, errors, and misunderstandings become apparent at the interfaces, and any evaluation effort must strive to define measures of effectiveness that sense or measure these disruptions. An effective aid is one that increases the speed with which information can be transmitted across these interfaces while decreasing the errors and misinterpretations on the part of the recipients of the information.

It is useful to think in terms of three of the many interfaces within a decision-making organization:

- the decision aid and user interface,
- the user and decision-making organization interface, and
- the decision-making organization and environment interface.

These interfaces identify three types of evaluation questions:

- Is the aid easy to use?

- Is the aid acceptable to the organization?

- Does the aid improve performance of the organization in relation to its environment?

These questions are ordered in terms of their difficulty of evaluation, with aid/user questions being most amenable to experimentation and organization/environment questions being least open to evaluation. This ordering arises as a result of the level of fidelity required for each type of evaluation.

Aid/user evaluations can tolerate low fidelity, thereby decreasing costs and permitting a higher level of experimenter control. Organization/environment evaluations require high fidelity, which both increases costs and permits much less experimenter control. User/organization evaluations fall between these extremes.

For each type of evaluation three methods of collecting measures of effectiveness are available: objective measurement, subjective judgment, and expert observation. Each method is best suited to a different notion of effectiveness. Objective measures are best suited to evaluating efficiency; subjective judgments are best suited to evaluating likability, acceptability, and tolerability; and expert observations are best suited to evaluating the correctness of a solution or inference. This association of collection techniques and concepts of effectiveness is simply a guideline and should not be interpreted as precluding the use of a technique to evaluate a type of effectiveness with which it is not associated.

A final question concerns what the evaluator intends to compare. In some rare instances, it could be appropriate to evaluate an aid in relation to some absolute scale and remain unconcerned about the aid's relation to other systems, but this is unlikely. Just as an experiment needs a baseline or control condition, so will an aid evaluation. The comparison could be between alternate designs of a single aid, between opposing aids, or between an aid and no aid, but in most instances, some comparison will be required. Otherwise, the evaluation results will lack a context and, therefore, a basis for deciding whether the aid is worthwhile.

These various elements of aid evaluation (i.e., measures of effectiveness, settings, and methods of collection) are basic to the problem of conducting an evaluation, but they do not represent all that is involved. Only through careful thought and effort can an evaluator pull these elements together for his or her specific problem. The purpose of this section has been to point the way. The hard work still remains.

REFERENCES

Ackoff, R. L. (1960). Unsuccessful case studies and why. Operations Research, 8, 259-263.

Allardyce, L. B., & Peterson, C. B. (1979). Hierarchical Evaluation Program (HIVAL) (Users Guide UG 79-2-67). McLean, VA: Decisions and Designs, Inc.

Allen, J. J., & Allardyce, L. B. (1978). Users manual for the software program MCCRESSA structure. McLean, VA: Decisions and Designs, Inc.

Amey, D. M., Feuerwerger, P. H., & Gulick, R. M. (1979a). Documentation of decision-aiding software: INFER users manual. McLean, VA: Decisions and Designs, Inc.

Amey, D. M., Feuerwerger, P. H., & Gulick, R. M. (1979b). Documentation of decision-aiding software: OPINT users manual. McLean, VA: Decisions and Designs, Inc.

Argyris, C. (1971). Management information systems: The challenge to rationality and enrationality. Management Science, 17, B275-B292.

Carter, E. E. (1972). What are the risks of risk analysis. Harvard Business Review.

Drake, J. W. (1971). The administration of transportation modeling projects. Lexington, MA: D. C. Heath.

Edwards, W. (1977). How to use multi-attribute utility measurement for social decision making. IEEE Transactions on Systems, Man, and Cybernetics, SMC-7, 326-340.

Ginzberg, M. J. (1978). Steps toward more effective implementation of MS and MIS. Interfaces, 8, 57-63.

Grayson, C. J., Jr. (1973). Management science and business practice. Harvard Business Review, 51, 41-48.

Gulick, R. M., & Allardyce, L. B. (1979). Documentation of decision-aiding software: R-SCREEN users manual (Users Manual 79-3-99). McLean, VA: Decisions and Designs, Inc.

Hammond, K. R., Cook, R. L., & Adelman, L. (1977). POLICY: An aid for decision making and international communication. The Colombia Journal of World Business, 12, 79-93.

Hammond, K. R., McClelland, G. H., & Mumpower, J. (1980). Human judgment and decision-making: Theories, methods, and procedures. New York: Hemisphere/Praeger.

Kaplan, M. F., & Schwartz, S. (Eds.) (1977). Human judgment and decision processes in applied settings. New York: Academic Press.

Keeney, R. L., & Raiffa, H. (1976). Decisions with multiple objectives. New York: Wiley.

Kelly, C. W. (1979). Program completion report: Advanced decision technology program (1972-1979) (Technical Report TR 79-3-93). McLean, VA: Decisions and Designs, Inc.

Lonnstedt, L. (1975). Factors related to the implementation of operations research solutions. Interfaces, 5, 23-30.

Powers, R. F., & Dickson, G. W. (1973). MIS project management: Myths, opinions, and reality. California Management Review, 15, 147-156.

Raiffa, H. (1968). Decision analysis. Reading, MA: Addison-Wesley.

Rubenstein, A. H., Radnor, M., Baker, N. R., Heiman, D. R., & McColly, J. B. (1967). Some organizational factors relating to the effectiveness of management science groups in industry. Management Science, 13, B508-B518.

Sage, A. P., & White, C. C., III (1980). Evaluation of two DDI decision aids developed for DCA:C140 (Document Number 33737-W114-RU-00). Falls Church, VA: TRW Defense and Space Systems Group.

Savage, L. J. (1959). The foundations of statistics. New York: Wiley.

Shycon, H. N. (1977). All around the model--Perspectives on MS applications. Interfaces, 7, 40-43.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. Annual Review of Psychology, 28, 1-39.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6, 649-744.

von Neumann, J., & Morgenstern, O. (1947). Theory of games and economic behavior (2nd ed.). Princeton, NJ: Princeton University Press.

Von Winterfeldt (1979).

Wolek, F. W. (1975). Implementation and the process of adopting managerial technology. Interfaces, 5, 38-46.