February 1981

TECHNICAL REPORT TR 81-1-304

Issues in the Design and Evaluation of Decision-Analytic Aids

Leonard Adelman
Michael L. Donnell
John F. Patterson
Jonathan J. Weiss

DECISIONS and DESIGNS, INC.

Prepared for
U.S. Army Research Institute
under Contract MDA903-80-C-0194
DARPA Order No. 3831

TECHNICAL REPORT TR 81-1-304

ISSUES IN THE DESIGN AND EVALUATION OF DECISION-ANALYTIC AIDS

by

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

Prepared for

U.S. Army Research Institute

Sponsored by

Defense Advanced Research Projects Agency
Contract MDA903-80-C-0194
DARPA Order No. 3831

February 1981

DECISIONS and DESIGNS, INC.
Suite 600, 8400 Westpark Drive
P.O.Box 907
McLean, Virginia 22101
(703) 821-2828

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR-81-1-304
2. GOVT ACCESSION NO.: AD-A096 893
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, Michael L. Donnell, John F. Patterson, Jonathan J. Weiss
8. CONTRACT OR GRANT NUMBER(s): MDA903-80-C-0194, DARPA Order-3831
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: DARPA Order No. 3831
11. CONTROLLING OFFICE NAME AND ADDRESS: Defense Advanced Research Projects Agency (DARPA) Cybernetics Technology Division, Defense Sciences Office, 1400 Wilson Blvd., Arlington, VA 22209
12. REPORT DATE: February 1981
13. NUMBER OF PAGES: 70 (74)
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 was prepared for the U.S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333.
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)

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 a consistent, realistic solution. 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. Decision-analytic aids

→ next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

390664

cont

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. This 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 of which both are a part. 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 or not 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 them.

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 of concern in the development and evaluation of decision aids that have arisen in the work of decision analysts at Decisions and Designs, Inc. (DDI) and elsewhere over the last few years. Such information should assist developers in integrating decision aids into their organization and, in turn, result in improved organizational decision making.

SUMMARY

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 a consistent, realistic solution. 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-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, evaluation, and implementation 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 of which both are a part. 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 or not 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 them.

ACKNOWLEDGMENTS

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, and Dr. Leonard Adelman represented DDI. The authors wish to thank all participants for contributions made during the conference and the Army Research Institute for supporting the colloquium under DSSW Contract No. MDA903-79-C-0275.

CONTENTS

CONTENTS (Continued)

FIGURES

TABLES

ISSUES IN THE DESIGN AND EVALUATION OF DECISION-ANALYTIC AIDS

1.0 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 a consistent, realistic solution. 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 Sections 2.0, 3.0, and 4.0 in this report is to provide guidelines for the effective design, implementation, and evaluation of such decision aids.

Throughout the following sections, the term "decision aid" or "decision-analytic aid" refers to a computer which has been programmed 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 which 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, manpower, etc.), compared to 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 it either improves or facilitates decision making, a device cannot properly be termed a decision aid.

Figure 1-1 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 within Figure 1-1. 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 of which both are a part. 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 or not 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.

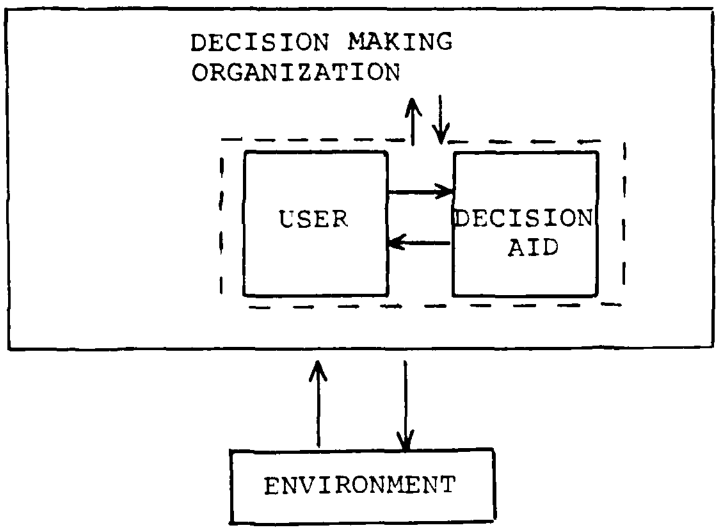

Figure 1-1

FRAMEWORK FOR CONSIDERING ISSUES RELEVANT
TO THE DESIGN AND EVALUATION OF DECISION AIDS

Section 2.0 addresses the first interface; starting with this 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 which 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 those 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 for them are offered.

Section 3.0 addresses the second interface in the framework, which considers how the decision aid is integrated into the larger decision-making organization. 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. The general behavioral characteristics of decision aids (the first interface) represent a necessary, but not sufficient condition for aid implementation. In addition, user involvement in aid design is essential for implementation, for it 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. Support for this position is presented from two sources: (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 indicate that it is often neglected in the development of analytical decision aids; the result is often an unsuccessful implementation effort.

In contrast to Sections 2.0 and 3.0, which focus on the design and implementation of decision aids, Section 4.0 considers the issue of 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 setting for decision aid evaluations. Evaluation settings differ in their similarity to the expected operational setting, the amount of control, 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 of 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 of concern in the development and evaluation of decision aids that have arisen in the work of decision analysts at Decisions and Designs, Inc. (DDI) and elsewhere over the last few years. Such information should assist developers in integrating decision aids into their organization and, in turn, result in improved organizational decision making.

2.0 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: that immediate, certain, directly observable effects have high impacts on behavior, while 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, on the other hand, 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, while only indirect feedback reaches the decision maker. Thus, the immediate process-related cost factors dominate the behavioral environment of decision aiding, while 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 which 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 terminate the analysis or, even worse, lead to a half-hearted attempt which 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 motivation 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. This requires the analyst to be more than just technically proficient; it requires a sensitivity to the client's mental and emotional state, and 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, to explain the procedures more carefully, to repeat a portion of the analysis, to call for a break, or simply to 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.

If it is a major challenge for the professionally trained decision analyst to maintain motivation and attention consistently, consider the even greater challenge to the decision aid which must accomplish the same task without the ability to observe the user's emotional and mental state, without the capacity to infer and adapt to the user's unique personal characteristics, and without the professional decision analyst's 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 results. Commercially produced video games* 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 behavioral-engineering features which 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 their eventual success.

The major contrast between the current generation of decision aids and the far more successful video games is that the former depends in a large part on the user's ability to endure unpleasant work in order to achieve a worthwhile long-range result, while the latter focuses on more

---

*"Pong," "Tank War," "Space Invaders," and so forth.

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.

2.1 Training, Initial Access, and Startup

Probably the most critical hurdle for decision aid usage is to get the user to sit down in front of the machine, turn it on, and start working. At that 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. While this phase is intrinsically an investment of sorts, every effort must be taken to minimize the effort and stress involved, while providing as much assistance and encouragement as possible. In particular, the following goals can help to overcome the initial block to usage.

2.1.1 Clear product identification - A decision aid should come with a clear "label" designating its intended use, the time and training it requires, its products, and a reference to a further source of information for potential users. 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.

2.1.2 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 on the area the user will be occupying.

2.1.3 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.

2.1.4 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.

2.1.5 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 importantly, 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 importance of user involvement in aid design to successful aid implementation is considered in detail by Adelman in the next paper in this volume.

2.2 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 ongoing dialogue, the aid's outputs should be clear and useful, while 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.

2.2.1 Simple controls - One of the most frequent complaints of potential decision aid users (particularly high-level decision makers) is the requirement to type all of their inputs on a keyboard. While an experienced typist or computer user might feel comfortable with a keyboard-based input method, a less experienced typist might experience impatience due to limited typing speed, frustration due to 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 decision aid 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 importantly, would eliminate the direct private and personal involvement of the decision maker in the process. Therefore, the use of a clerical "assistant" must be regarded as unacceptable in most cases.

Fortunately, some simple alternative approaches can eliminate the need for most keyboard inputs. These may range from the simple buttons and levers which control commercial computer games through futuristic devices such as speech recognizers, visual pattern detectors, and pressure-sensitive display screens. While further human-engineering efforts will be needed to determine the most effective functions, sizes, positions, and configurations, these simplified input devices can already produce quantum 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.

2.2.2 Continual sensorimotor feedback - When two people interact in normal conversation, the information transmitted is more than just the words which are being uttered. The "carrier wave" which makes a conversation seem natural is an almost unnoticed background of eye motions, nonverbal noises, postural adjustments, and facial mechanisms, 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 to control it, the conversation would become awkward and its participants would feel uncomfortable. Occasional long pauses would occur, making the participants unsure about whether the dialogue should continue. On the other hand, sometimes two or more parties would attempt to speak at once, inducing not only confusion but also a certain amount of social friction as well. The nonverbal component of a face-to-face conversation avoids these problems, smoothing the social content and assuring efficient communication.

The same observation applies to the "social" 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." First, the aid must provide continuous or frequent indication that all functions are working properly; for example, a clock whose display changes every second will assure the user that the aid is running satisfactorily. Next, the aid must provide immediate sensorimotor feedback by acknowledging every input without perceptible delay using 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 (when experienced users have previously been "spoiled" into expecting 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 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.

2.2.3 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 should direct the user's attention to the proper information, require little effort to "read" that 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 by pictorial symbols, photographic images, graphs, maps, and so forth.
- 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 in displays 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.

2.2.4 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 still exceed the user's limits. A human factors engineering approach 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 errors. Altering a display configuration or an input device may have a significant impact on the amount of strain imposed, and considerations of 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.

2.2.5 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 which 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 which might arise from ambiguous decision-analytic ones such as "utility," "attribute," "risk," "weight," and "option." Where 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.

2.3 Intermediate Reinforcement to Increase Attention and Motivation

The preceding sections have covered ways to attract the user to the aid and to simplify its overall operation. 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 towards the ultimate goal. In addition, it must help the user to follow that path without losing sight of the goal or getting distracted along the way. Because the successful use of a decision aid is 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 itself 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 completion of intermediate goals;
- timing and sequencing tasks to avoid boredom, and
- reinforcing the user for maintaining a high level of attention.

2.3.1 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 towards 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.

2.3.2 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 (using different sensory modalities, different display colors, different muscular movements, etc.), 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.

2.3.3 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 permited to clear his 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 programmed 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 and used 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 (See Section 3.0 in this report) or to adapt procedures to individual needs (See Section 2.5 of this paper) will be especially important.

2.4 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 "business" part of the reinforcement might include the following:

- hard copies of tables, charts, graphs, etc., which might be useful in briefing the outcome of the analysis to others (or as input into some higher-order decision process);

- a preformatted report which 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 importantly, the aid should acknowledge the user's hard work, and elicit any comments, suggestions, or questions which might help to improve future versions.

2.5 Adapting Procedures to Meet Individual Needs

If the suggestions specified in Sections 2.1 and 2.2 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 several individual users' behavior rather than on a single group decision, the ability to fine-tune the aid to individual specifications will dramatically improve its acceptance. The more variability among individual users (or individual problems), the more important this customization will become.

2.5.1 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 where 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 so forth. 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 which influence the aid's availability and usage, and from the urgency of other tasks competing for the decision makers' time.

2.5.2 Adaptability to various training levels - One universal problem with respect to 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. On the other hand, as the aid becomes 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 with 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 that the benefits

of the machine's "hand holding" are no longer worth the extra time required, a more streamlined version 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 tradeoffs, etc.) 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 still more sophisticated version of this capability might keep a record of a decision maker's past usage, error rates, speed of response and so forth, automatically starting the user at the most appropriate level and modifying the level based on current performance (but always subject to user override).

2.5.3 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 items such as the choice of a color-coding scheme, symbology, and display formatting may initially assume some arbitrary default setting, but on the user's request, may 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, where 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?", etc.) 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 provide all users with an automatic capability to adapt some parameters based on its observation of the user's "behavioral state" inferred from response times, error rates, answers to direct inquiries (e.g., "Do you want to continue or would you like a break?") and, if available, physiological monitoring. 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.

2.6 Conclusion

In the future, computer-aided decision making may bear little resemblance to those methods available today. As the general public becomes more knowledgeable about computers, and as computer usage by nonspecialists becomes widespread, some of the blocks which 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 Section 2.0, 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.

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

Over the last twenty-five 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 advances in the development of normative decision theory (e.g., von Neuman and Morgenstern, 1947; Savage, 1954; Raiffa, 1968; Keeney and 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 include different types of (a) multi-attribute utility assessment programs such as HIVAL (Allardyce and Peterson, 1979), and POLICY (Hammond, Cook, and Adelman, 1977), as well as (b) traditional decision-analytic aids requiring probability and utility assessment, such as INFER (Amey, Feuerwerger, and Gulick, 1979a) and OPINT (Amey, Feuerwerger, and 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 Kelly (1978).

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.

---

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

The thesis of Section 3.0 is that decision-analytic aids will seldom achieve a stand-alone status unless eventual users are involved in their development. The term "users" applies here both to the persons running the decision aid and to the decision makers utilizing its results. In the previous section, we focused on the interface between the aid and its "hands-on" users, who may or may not be decision makers. The position, simply put, 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 of its successful integration into the organization. In this section, the focus is on interface between the user (and decision aid) and the larger decision-making organization of which both are a part. The position here is that the general behavioral characteristics of decision aids represent a necessary, but not sufficient condition for aid implementation. In addition, involvement of decision makers in aid design is essential for implementation, for it 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 importance of user involvement in aid design to successful aid implementation has not been emphasized in the judgment/decision research literature. Support for this position, therefore, 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 is is often neglected in the development of analytical decision aids; the result is often an unsuccessful implementation effort.

3.1 Evaluation of Decision-Analytic Aids

3.1.1 R-SCREEN - Sage and White (1979) recently evaluated a multi-attribute utility assessment (MAUA) aid, called R-SCREEN (Rapid Screening of Decision Options), developed by DDI under funding provided by the Command and Control Technical Center, Defense Communications Agency (DCA-C140). 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 (JCS).

DCA-C140 is tasked by DCA with monitoring, evaluating, and improving the overall information flow within the World Wide Military Command and Control System (WWMCCS) in support of information reporting, information analysis, decision making, and information dissemination. Consistent with this focus, DCA-C140 recently conducted an experiment concerned with exploring the usefulness of computer-based decision aids in improving the option generation and selection process as it occurs in command centers in crisis situations. In support of this effort, DDI was tasked to modify an existing generic decision aid previously used by its analysts in an operational setting with the European Command (EUCOM).

JOD decision makers used R-SCREEN by implementing four steps. First, they selected one of the three prestructured templates (or hierarchies) 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'[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. DDI analysts briefed the JOD staff on how to use the aid, developed a user's guide specifically for the aid (Gulick and 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 six months.

Sage and White evaluated R-SCREEN by the following three procedures: (1) informal interviews with JOD personnel and others familiar with the JOD operational environment; (2) 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 three criteria categories and subcriteria shown in Table 3-1. Although Sage and White discussed the implications of the responses in terms of each of the fifteen subcriteria, they did not give R-SCREEN an explicit score on each criterion. In order to shorten this presentation, the present authors gave R-SCREEN an overall score (+, -, or ?) based on Sage and White's qualitative evaluation on each subcriterion.

R-SCREEN rated extremely well in terms of performance objective achievement (category 1) and efficacy (category 3).

Algorithmic Effectiveness or Performance Objective Achievement Evaluation

. logical soundness	+*
. improved decision quality	+
. decision process changes	+

Behavioral or Human Factors Evaluation

. political acceptability	-
. institutional constraints	-
. implementability	?
. procedural changes	-
. side effects	?

Efficacy Evaluation

. time requirements	-
. leadership and training requirements	?
. communication accomplishments	?
. educational accomplishments	+
. documentation	+
. reliability	+
. convenience of access	+

*Notation:

+ means performed well

- means performed poorly

? means rated well and poorly on questions comprising the subcriterion category

Table 3-1

CRITERIA FOR DECISION AID EVALUATION AND EVALUATION SCORES (+, -, or ?) FOR R-SCREEN

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 problem, and facilitate communication (e.g., see Hammond et al., 1980).

R-SCREEN rated poorly, however, in terms of 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. On these two points, Sage and White are quoted at length below:

> "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 match 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." (1979, p. 0.11)

In short, R-SCREEN was not tailored to the personal needs and organizational context of its eventual users. As a result, evaluation responses indicate that its implementation into JOD was questionable--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 that 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 below in an effort to help readers evaluate the adequacy of the thesis advanced in this section.

3.1.2 MCCRESSA - The U.S. Marine Corps (USMC), as well as other services, has a continuing problem in assessing, under peacetime conditions, the combat readiness of their combat units. The problem is compounded by the many heterogeneous attributes that are used to describe the performance of individual combat units and 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 combat readiness evaluation process or validation of its results.

DDI was tasked by the Defense Advanced Projects Agency (DARPA) to develop a decision aid to support the Marine Corps Combat Readiness Evaluation System (MCCRES). DDI developed an MAUA aid called MCCRESSA, for Marine Corps Combat Readiness Evaluation System Software Application (Allen and Allardyce, 1978). It was successfully tested in an operational setting by the Marine Corps in August 1977 and is now in routine use throughout the Marine Corps.

MCCRESSA and R-SCREEN are extremely similar. Both are MAUA decision aids. Both were designed from the same generic MAUA software used by DDI analysts. If anything, R-SCREEN was more sophisticated analytically than MCCRESSA because it forced the user to assign criterion weights moving from the bottom to the top of the hierarchy, thereby ensuring that the upper-level weights were determined by the scores on the lower-level attributes and not by the user's general perception as to 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 development and implementation process. DDI analysts worked directly with the five Marine colonels tasked with developing and implementing MCCRES over a one-year period. These men decided on the criteria, hierarchial 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 output 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 two-day conference with 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 it 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 the development and implementation process ensured that MCCRESSA was tailored to the Marine Corps' needs and organizational context. We believe this to be the principal factor in MCCRESSA's successful implementation into the USMC.

3.2 Implementing Operations Research Models

Although there are distinct differences between decision-analytic aids and operations research (OR) 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 last two decades has (a) documented numerous cases where analytically rigorous OR models go unused by the client for whom they were developed, (b) tried to explain this phenomenon, and (c) offered suggestions for how to minimize it. This section briefly reviews this literature.

Ginsberg (1978) divides the OR/MS literature on implementation into two types: the normative approach (e.g., Ackoff, 1960; Argyris, 1971; Grayson and Jackson, 1973) and the factor study (e.g., Drake, 1973; Powers and Dickson, 1973; Rubenstein et al., 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* [italics ours]." (Ginsberg, 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 twelve 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 which were not implemented by the user.

The results of the survey are presented in Table 3-2. As can be seen, there is a positive relationship between implementation and (a) the user's collaboration in defining the problem, (b) problem characteristics, and (c) 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 course of the model development process.

Shycon (1977) conducted two surveys of large OR projects varying in their degree of successful implementation.

VARIABLE GROUP	VARIABLE	CHI SQUARE	SIGNIFICANCE (P=)	INFLUENCE OF NON RESPONSES 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

Table 3-2

SUMMARY OF RESULTS: RELATIONSHIPS BETWEEN VARIABLES AND PROPOSAL IMPLEMENTATION (From Lonnstedt, 1975)

Both surveys categorized each OR project team's organization according to 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. Of these three types, the decision-analytic project team's organization for implementing R-SCREEN best represented the wholly management science team. In contrast, the team's organization for MCCRESSA best represented the interparticipative M/S team.

Table 3-3 presents the results of Shycon's (1977) surveys. The wholly management science team achieved the lowest degree of implementation success. On the other hand, the M/S team with communication and the interparticipative M/S team consistently achieved a high level of implementation success. Neither project team organization types 2 nor 3, however, showed any distinct advantage over the other. Given the greater cost of category 3, the results suggest that the M/S team with communication is the most cost-effective project team organization. Nevertheless, the results support post hoc the obtained outcomes for R-SCREEN and MCCRESSA regarding implementation success.

3.3 Why User Involvement is Essential

User involvement throughout the development of decision-analytic aids is essential to (a) making users comfortable with a decision-analytic approach to decision making, and (b) learning enough about the user's goals and working

PROJECT TEAM ORGANIZATION AND PROJECT SUCCESS (1971)

TYPE OF PROJECT	TYPE OF FIRM	APPROX. PROJECT COST[1] $000	DEGREE OF SUCCESS
1. The Wholly Management Science Team			
Determination of regional distribution requirements	Regional food distribution	36	Partial
Sales forecasting & inventory planning system	National tool manufacturer	68	Partial, long-term
Sales forecasting & inventory planning system	Division of major drug manufacturer	36	Little immediate, some longer-term
Determination of service call response strategy & facilities required	Regional public utility	30	Completely successful
Determination of service call response strategy & facilities required	Regional public utility (different from above)	32	Partial
2. The MIS Team with Marginal Communication			
Design of national distribution system	National food processor	140	Completely successful
Design of total logistics system, manufacturing plants and distribution	National food processor	95	Completely successful
Service facility requirements	Heavy machinery manufacturer	55	Completely successful
Integrated distribution requirements for diverse divisions	Major drug & toiletries manufacturer	54	Minor benefits
3. The Fully Interparticipative MIS Team - Participation and Communication at All Levels			
Total management planning program: procurement, inventory, scheduling, distribution sales forecasting & marketing planning	Major national meat packer	225	Completely successful
Design of national distribution system	National food processor	120	Completely successful
Design of national distribution system	Major instrument and supplier manufacturer	96	Completely successful
Basic simulation of company operations for management policy testing	Major pharmaceutical manufacturer	104	Completely successful
Evaluation system for R&D projects & tool program for rank order and funding	Major synthetic fiber	160	Partial success
Corporate strategy model for decisions in marketing, manufacturing, capital investment	Major integrated paper products manufacturer	250	Completely successful
Design of total logistics system, manufacturing plants and distribution	National industrial products plastics manufacturer	75	Completely successful

[1] Includes all services, internal personnel, external expenses incurred. All figures adjusted to 1971 dollars.

PROJECT TEAM ORGANIZATION AND PROJECT SUCCESS (1977)

TYPE OF PROJECT	TYPE OF FIRM	APPROX. PROJECT COST[1] $000	DEGREE OF SUCCESS[2]
1. The Wholly Management Science Team			
Determination of national distribution and transportation strategy	Consumer large instruments	42	Little
Multi-plant manufacturing operational strategy	Building products manufacturer	77	Little
2. The M/S Team with Communication			
Research and development of inventory planning system	Sporting goods manufacturer	110	Complete
Development of marketing strategy with attendant distribution requirements	Automobile after market manufacturer	85	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. The Interparticipative M/S 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 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

[1] Includes all services, internal personnel, external expenses incurred. All figures adjusted to 1977 dollars.

[2] Degree of success is necessarily partly subjective, however measurable criteria are: 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.

Table 3-3

PROJECT TEAM ORGANIZATION AND PROJECT SUCCESS (From Shycon, 1977)

environment so that the decision-analytic aid can be tailored to their personal needs and organizational context.

Decision makers are not decision analysts. While 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 analyst 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 their 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 they 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 the aid will be used permits analysts to tailor decision-analytic aids 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 successful and unsuccessful 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 as R-SCREEN illustrated, just because a decision-analytic aid is recognized as beneficial does not mean that it will be adopted; the aid also has to score well on behavioral characteristics.

Decision-analytic aids will change the users' working environment, for they 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 one with a previously determined structure and stored data (e.g., weights) like MCCRESSA (2) a multi-purpose one with no substantive structure and stored data, or (3) an aid like R-SCREEN that combines features of both extremes. What will vary is the extent of the change and its effect on the interpersonal relations and functions of different people within the organization. In addition, it may be necessary to hire (or train internally) skilled decision analysts to use effectively multi-purpose 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, i.e., "...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 thesis of Section 3.0 may not surprise many decision analysts and researchers; the need for user involvement in aid development seems obvious. What may surprise them is the empirical support for it; 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 its importance 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 Section 3.0 may well lie in alerting analysts to the necessity of making users realize the importance of their involvement in aid development to successful implementation.

4.0 DECISION AID EVALUATION

The preceding sections have been concerned with the problem of designing effective decision aids (DAs), while the present section turns to the problem of evaluating DAs. 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 it 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 evaluation questions relate to the evaluation of decision aids; these questions correspond to the three types of interfaces identified in the introduction to this report (Figures 1-1 and 4-1). First, an evaluation may attempt to answer questions about the DA's compatibility with its immediate users (DA/U). Such questions are concerned with the human factors of the DA, for example: Are its displays effective? Is it tedious to use? In addition, the DA/U interface is the point at which questions about the comprehensibility of a DA's underlying model are addressed.

A second type of evaluation occurs at the interface between a DA user (and the decision aid) and the remainder of the decision-making organization (DMO). The U/DMO interface poses questions about the collectibility of a DA's

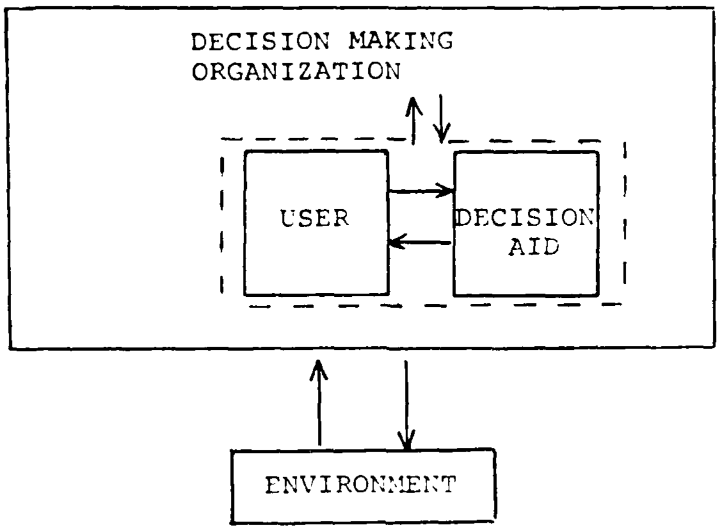

Figure 4-1

FRAMEWORK FOR CONSIDERING ISSUES RELEVANT
TO THE DESIGN AND EVALUATION OF DECISION AIDS

required inputs and the communicability of a DA's outputs. A DA which is only comprehensible to its immediate user is likely to be useless in most DMOs. 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 DMO and its environment (DMO/ENV). At this point the ultimate question of a DA's effectiveness comes into play; namely, has the DA improved the DMO's output or performance. Similarly, there are questions about the range of environments or problem areas over which the DA provides improved DMO performance.

These three types of interfaces--DA/U, U/DMO, and DMO/ENV--are by no means independent. In fact, they are "nested": U/DMO effectiveness is necessarily influenced by DA/U effectiveness, and DMO/ENV 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 DA evaluation.

The sections that follow use this DA evaluation framework to examine several aspects of evaluation. Section 4.1 considers the problem of identifying measures of effectiveness (MOEs). Section 4.2 addresses the selection of a setting for conducting the evaluation. Section 4.3 discusses the selection of a method for data collection. And Section 4.4 discusses the question of what is being compared in the evaluation.

4.1 Measures of Effectiveness (MOEs)

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 many measures of effectiveness (MOEs), which are designed to answer the evaluator's questions. Ideally, these MOE's are objectively measurable and quantitative variables that will describe the effectiveness of a DA. In the present case, however, the term MOE will also include subjectively measurable variables and variables that result in a qualitative rather than quantitative description. The only restrictions are that the MOE must be measurable and that it should be expected to correlate (positively or negatively) with the effectiveness or efficiency of a DA.

While it would be impossible to list the MOEs that are appropriate to all evaluations, another approach may provide some insight into MOEs. Consider the process by which a DA is used to be divisible into stages which proceed from initial data collection to decision implementation (see Figure 4-2). Under this assumption, anything that improves the effectiveness or efficiency of one stage should improve the overall effectiveness. Thus, the discussion of MOEs can be simplified somewhat to a discussion of how the DA can affect each stage.

4.1.1 Data collection - Data collection is the stage during which the DMO 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 DMO/ENV interface.

At first glance, this stage might seem remote from the DA and therefore not subject to its influence. Nothing could be further from the truth. The introduction of a DA has the potential to greatly improve this stage by demanding a better organized and better quantified measurement of the environment. In addition, it presents the

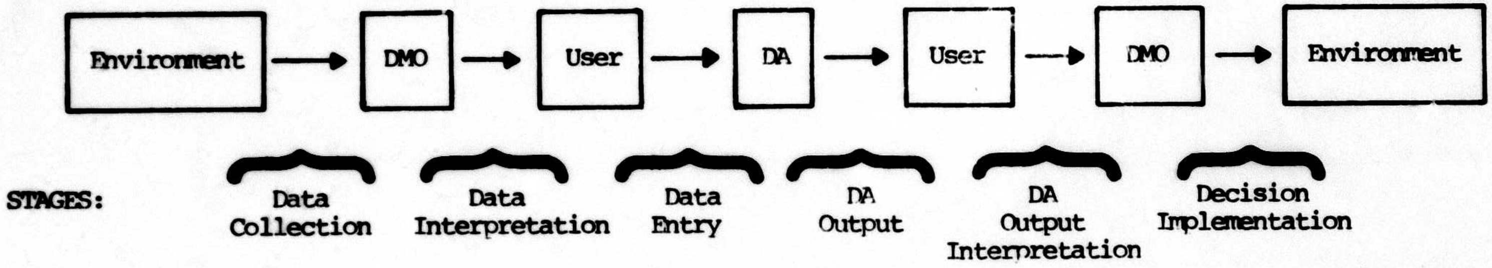

Figure 4-2
STAGE MODEL OF DA USAGE

possibility of directly interfacing the sensing equipment with the DA which could, in some instances, eliminate errors that might otherwise be introduced as the data is transferred through the DMO to the user and into the DA.

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

4.1.2 Data interpretation - Strictly speaking, the data interpretation stage is the stage during which the members of the DMO 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 DMO, it is considered here as part of the U/DMO interface, since this is the point at which the final judgments must be made.

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

In contrast to these benefits, the DA may introduce substantial costs to the data interpretation stage. It 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 on the DMO by requiring its members to think in a manner that is neither natural nor intuitive.

4.1.3 Data entry - The data entry stage is the period during which the user provides the DA with its required inputs. This stage is part of the DA/U interface.

Data entry is a necessary and frequently arduous aspect of DA usage. For this stage a DA 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 in the first section of this report.

4.1.4 DA output - The DA output stage is the period during which an aid provides a user with its results. As such, it is clearly part of the DA/U interface. The user for this stage need not, however, be the same individual as the user in the data entry stage.

During the data output stage the DA has its most obvious and immediate opportunity to be of value to the DMO. To accomplish this it 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. Rapid editing may be required, lest the user discover inaccurate or irrational judgments. And, 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 a DA will not be without cost. Besides the equipment itself, DA 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 DA's underlying model and the options available for its exploration can pay off. The job of conveying the model's results to the DMO will fall upon this user.

4.1.5 DA output interpretation - The DA output interpretation stage is the period during which the implications of the DA's analysis are conveyed to the DMO and a decision is made. Although this stage involves many actors in the DMO, and may only briefly involve the DA user, it is considered, nonetheless, as part of the U/DMO interface.

The DA's influence on this stage will depend on its 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 DMO, then its results have a chance of acceptance. Especially important are the communicability and justifiability of the DA's implications. If, on the other hand, anyone in the chain of communication leading to the decision makers or the decision makers themselves feel annoyed at or uncertain about the DA's results, then these results are likely to be ignored and possibly suppressed from that point on. One way to minimize this occurrence is to have users (both "hands-on" users and decision makers) involved in designing the decision aid. By doing so, one 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 DMO.

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

For this stage the most fundamental question is whether e DA has led to a sound decision. If its implications are correct and they are not ignored, then the DA has provided its major benefit. It 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.

4.1.7 Summary of potential MOEs - To provide a summary of some potential MOEs, Figure 4-3 reiterates many of the issues raised in the preceding sections. In this figure, the issues are organized into a hierarchy, and it is assumed that each terminal node could be translated into an MOE. This representation is not meant to advocate any particular set of MOEs or any particular approach to evaluation. It is simply a summary of a number of MOEs that may be relevant to any specific DA evaluation.

4.1.8 Organizational impact - Although it is implicit in the comments of the preceding sections, one additional point deserves mention. The introduction of a DA into a complex organization is unlikely to be accomplished without changes in the organization. A DA 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 DA

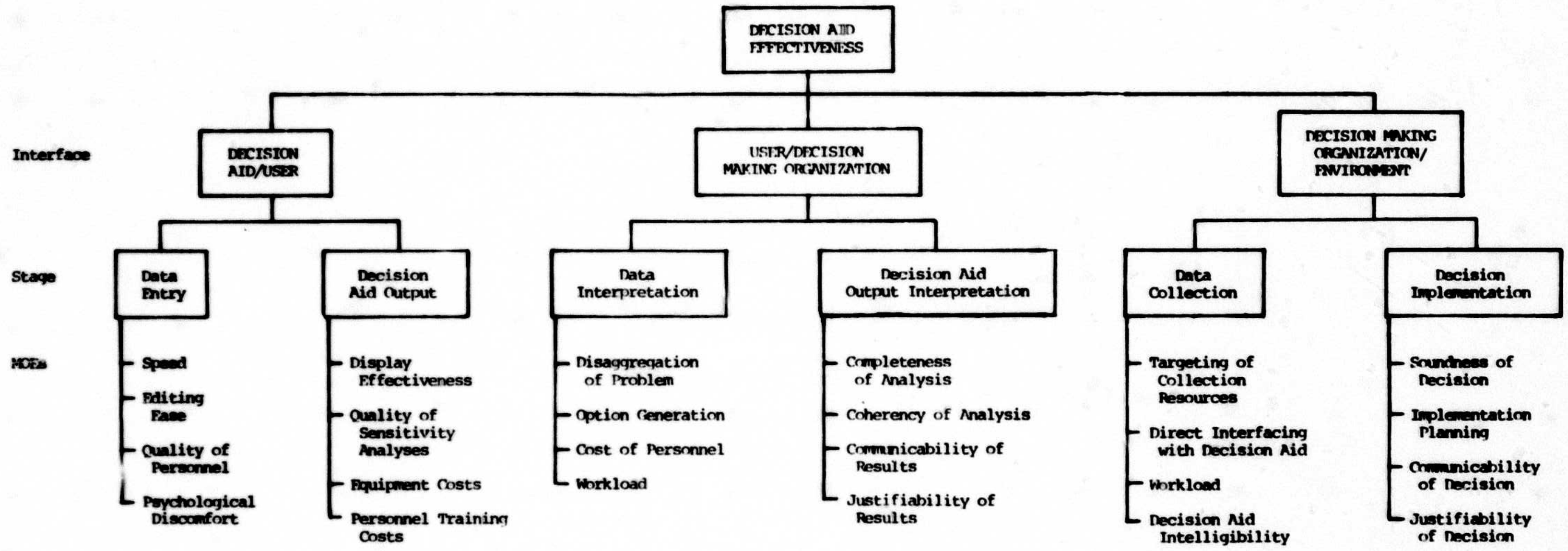

Figure 4-3

SUMMARY OF POTENTIAL MOEs AFFECTED BY DA will be unlike any existing component, i.e., individual, of the DMO. It cannot truly replace a person, because it cannot do all that a person does. This very fact implies that the introduction of the DA will necessarily redefine certain roles within the DMO.

The impact of a DA on an organization may in fact be sweeping. To use the DA, 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 DA. 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 DA will have a chance to take hold and contribute to the effectiveness of its DMO.

4.2 Settings for DA Evaluations

Before the MOEs for a DA can be collected and analyzed, it is necessary to construct a setting in which the DA 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 4-4 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 DA. Of course, this setting will not be available for evaluation purposes unless the DA is actually deployed. In lieu of the target setting, it is therefore necessary to construct a test setting within which the evaluation can proceed.

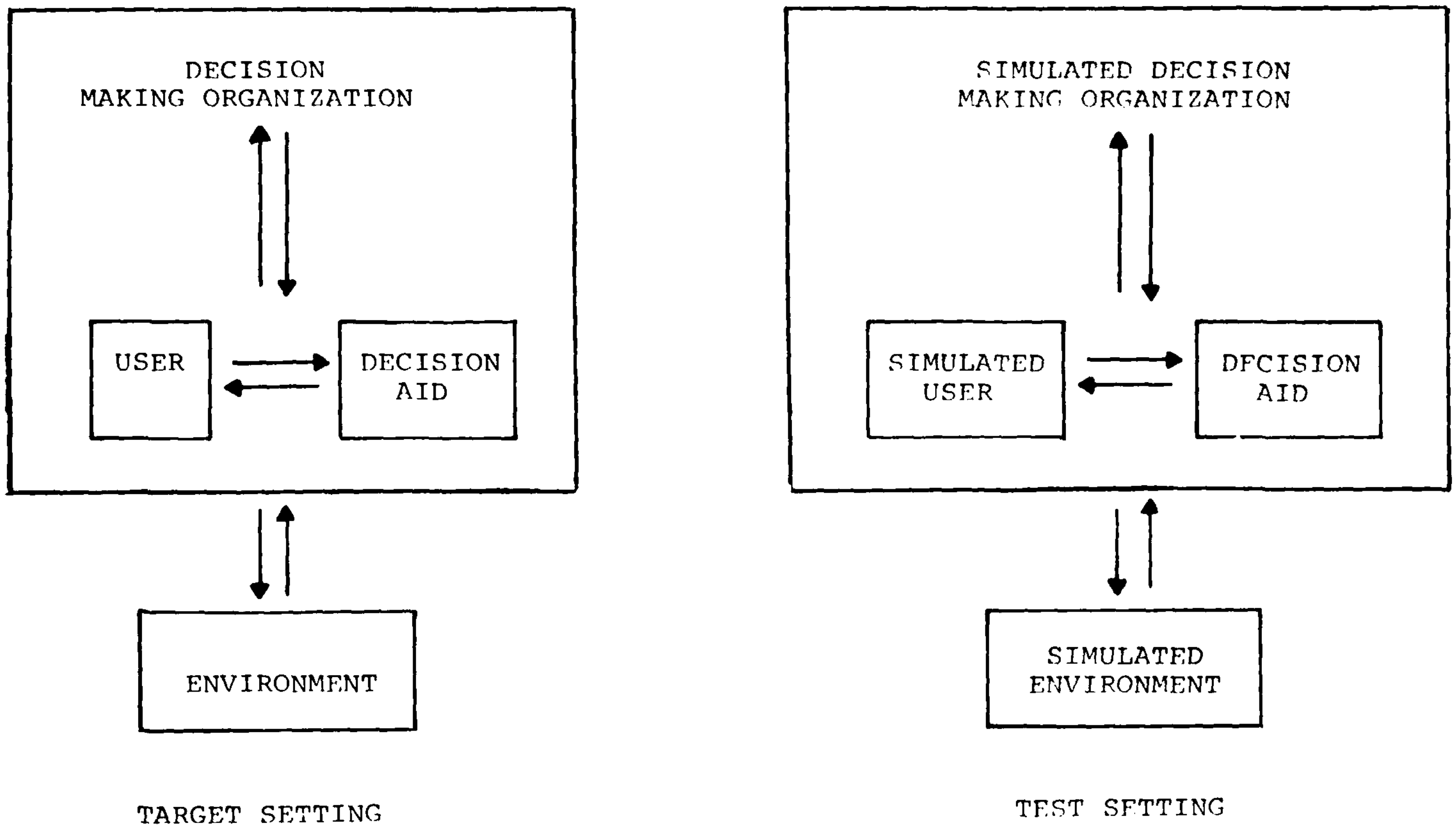

Figure 4-4

NOTIONAL REPRESENTATION OF THE SETTING FOR A DA EVALUATION

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 DMO, 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. Besides increased dollar costs and evaluation time, fidelity introduces an additional 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 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: those involving either high or low fidelity for the DMO and 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 low-fidelity environment and low-fidelity DMO is the most austere. Such a setting is well suited to DA design questions concerning its user compatibility. These questions are primarily concerned with the DA/U interface and therefore need not concern themselves with the DMO or even the true environment. Also, since so little simulation effort is required, this setting can be implemented in a carefully controlled laboratory experiment.

The high-fidelity environment and low-fidelity DMO setting can also be conducted in a laboratory setting, but it serves a different purpose. In this setting, great attention is paid to providing the user with realistic data about the environment, realistic options, and realistic scenarios, but the DMO through which the user would interact with this environment is only superficially implemented. Thus, it is possible to investigate the coherence and completeness of the model underlying the DA without going the additional step and ascertaining whether it will improve DMO performance. Such an analysis really only evaluates questions related to the DA/U interface.

The low-fidelity environment and high-fidelity DMO setting provides the means to answer questions about the U/DMO 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 DMO. Since a simulated DMO 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 U/DMO interface, it implies decreasing control and increasing costs.

Finally, the high-fidelity environment and high-fidelity DMO setting 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 DA effects on DMO performance. Since these questions about the DMO/ENV interface are the ultimate questions concerning DA effectiveness, field tests of this sort are a highly desirable precursor to DA deployment.

Figure 4-5 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. Thus, 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 DA coherence and completeness, questions of DMO compatibility, and questions of the DA effect on DMO 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.

4.3 Methods for MOE Collection

There are three major methods for obtaining MOEs: 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 observations, also involves subjective judgments, only this time 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.

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

Figure 4-5

SUMMARY OF ALTERNATIVE EVALUATION SETTINGS

In DA evaluation, objective measurements are likely to consist of speed and frequency measures. It will be important to know how long some process requires or how frequently errors occur. Figure 4-6 suggests ways in which these objective measurements could be used to evaluate the DA/U, U/DMO, and DMO/ENV interfaces.

No less important than objective measures are the assessments of a participant's satisfaction, complaints, or other judgments about a DA'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 a DA's ultimate acceptability. An aid which is objectively effective, yet subjectively unacceptable, is still unacceptable, since its chances for effective deployment are low. Thus, subjective judgments should not be ignored. Figure 4-6 suggests several MOEs which measure subjective judgments for each of the three interfaces.

The final method of evaluation, expert observations, differs from subjective judgments in that the raters are no longer 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 "truly" 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. They are deemed correct by virtue of their greater experience. Figure 4-6 suggests some ways that expert observations can assist an evaluation effort.

In summary, the three methods of MOE collection are: objective measurement, subjective judgment, and expert observation, each playing an important role with regard to

METHOD	DECISION AID/USER	USER/DECISION MAKING ORGANIZATION	DECISION MAKING ORGANIZATION ENVIRONMENT
Objective Measurement	Time for Data Entry	Time for Data Interpretation	Speed and Accuracy of Decision Implementation
Subjective Judgments	Rating of DA Display Effectiveness	Rating of Communica-bility of DA Outputs	Rating of Comprehensibility of Decision
Expert Observations	Rating of Completeness DA Analysis	Rating of Soundness of Data Interpretation	Rating of Soundness of Decision

Figure 4-6

MOEs FOR EACH METHOD AS APPLIED TO EACH INTERFACE 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, where such objective definitions are usually unavailable.

4.4 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 control condition. Similarly, some notion of a control condition or at least a contrasting condition is required for DA evaluations.

Consider the three interfaces once again. Evaluations involving the DA/U interface are largely concerned with user compatibility and DA coherence. As such it is reasonable to consider a comparison between alternate configurations of the DA or even between different DAs. A comparison between a DA and no DA is, however, inappropriate. Questions about the DA/U interface assume a DA just as they assume a user.

Evaluations involving the U/DMO or DMO/ENV interfaces permit more comparisons than the previous type of evaluation. In particular, it is now possible to examine the DMO operation both in the presence and in the absence of the DA. To perform such an evaluation one must recognize, however, that the definition of the user--and, therefore, the U/DMO interface--will change when the DA is not present. In other words, a DA is not like a plug-in module; its introduction to the DMO will necessarily require changes in the organization of the DMO. Thus, an effort to compare performance in the presence versus absence of the DA may be confounded with effects of the DMO changes required by the DA.

While effective performance is likely to be welcome regardless of whether it is caused by the DA or the changes that a DA requires, this confounding should not be overlooked. The scenario of primary concern is one in which the bulk of the benefit attributed to a DA is actually due to its concomitant DMO changes. If this possibility exists, a thorough examination requires that the evaluator attempt to compare performance in the modified DMO in the presence and in the absence of the DA. Under the assumption that the necessary DMO changes have been implemented, it should be easier to unplug the DA than it would have been to plug in the DA before the DMO changes.

4.5 Summary

DA evaluation begins with a recognition of the fact that the DA will simply be one component of a more complex information processing system. As such the inputs to the DA and the outputs from the DA will probably travel through many layers of a complex decision-making organization. Thus, a thorough analysis of DA effects is likely to examine aspects of the DMO performance as well as the DA 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 (MOEs) that sense or measure these disruptions. An effective DA 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.

Of the many interfaces within a DMO, it is useful to think in terms of three:

- the DA and user interface (DA/U);
- the user and DMO interface (U/DMO); and
- the DMO and environment interface (DMO/ENV).

These identify three types of evaluation questions:

- Is the DA easy to use?
- Is the DA acceptable to the DMO?
- Does the DA improve DMO performance in relation to its environment?

These questions are ordered in terms of their difficulty of evaluation, with DA/U questions being most amenable to experimentation and DMO/ENV questions being least open to evaluation. This ordering arises as a result of the level of "fidelity" required for each type of evaluation. DA/U evaluations can tolerate low fidelity thereby decreasing costs and permitting a higher level of experimenter control. DMO/ENV evaluations require high fidelity, which both increases costs and permits much less experimenter control. U/DMO evaluations fall between these extremes.

For each type of evaluation three methods of MOE collection are available: objective measurement, subjective judgment, and expert observation. Each of these 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 unassociated.

A final question, which a DA evaluator should confront, concerns what he intends to compare. In some rare instances, it could be appropriate to evaluate a DA in relation to some absolute scale and remain unconcerned about its relation to other systems, but this is unlikely. Just as an experiment needs a baseline or control condition, so will a DA evaluation. The comparison could be between alternate designs of a single DA, between opposing DAs, or between a DA and no DA 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 DA is worthwhile or not.

These various elements of DA evaluation, i.e., MOEs, settings, methods of collection, and the comparison, are basic to the problem of conducting an evaluation, but they do not represent its entirety. Only through careful thought and effort can an evaluator pull these elements together for his specific problem. The purpose of Section 4.0 has been to point the way. The hard work still remains.

REFERENCES

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

Allardyce, L. B.; and Peterson, C. B. Hierarchical Evaluation Program (HIVAL). Users Guide UG 79-2-87. McLean, Virginia: Decisions and Designs, Inc., 1979.

Allen, J. J.; and Allardyce, L. B. Users Manual for the Software Program MCCRESSA Structure. McLean, Virginia: Decisions and Designs, Inc., 1978.

Amey, D. M.; Feuerwerger, P. H.; and Gulick, R. M. Documentation of Decision-Aiding Software: INFER Users Manual. McLean, Virginia: Decisions and Designs, Inc., June 1979a.

Amey, D. M.; Feuerwerger, P. H.; and Gulick, R. M. Documentation of Decision-Aiding Software: OPINT Users Manual. McLean, Virginia: Decisions and Designs, Inc., April 1979b.

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

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

Drake, J. W. The Administration of Transportation Modeling Projects. Lexington, Mass.: D. C. Heath, 1971.

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

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

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

Gulick, R. M.; and Allardyce, L. B. Documentation of Decision-Aiding Software: R-SCREEN Users Manual. Users Manual 79-3-99. McLean, Virginia: Decisions and Designs, Inc., 1979.

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

Hammond, K. R.; McClelland, G. H.; and Mumpower, J. Human Judgment and Decision-making: Theories, Methods, and Procedures. New York: Hemisphere/Praeger, 1980, forthcoming.

Kaplan, M. F.; and Schwartz, S. (Eds.) Human Judgment and Decision Processes in Applied Settings. New York: Academic Press, 1977.

Keeney, R. L.; and Raiffa, H. Decisions with Multiple Objectives. New York: Wiley, 1976.

Kelly, C. W. Program Completion Report: Advanced Decision Technology Program (1972-1979). Technical Report TR 79-3-93. McLean, Virginia: Decisions and Designs, Inc., 1979.

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

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

Raiffa, H. Decision Analysis. Reading, Mass.: Addison-Wesley, 1968.

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

Sage, A. P.; and White, C. C., III. Evaluation of Two DDI Decision Aids Developed for DCA:C140. Document Number 33737-W114-RU-00. Falls Church, Virginia: TRW Defense and Space Systems Group, January 1980.

Savage, L. J. The Foundations of Statistics. New York: Wiley, 1954.

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

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

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

von Neumann, J.; and Morgenstern, O. Theory of Games and Economic Behavior. 2nd ed. Princeton, N.J.: Princeton University Press, 1947.

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