DOCUMENT RESUME

ED 360 346 TM 020 216

AUTHOR	Rouse, William B.; Johnson, William B.
TITLE	Computational Approaches for Analyzing Tradeoffs between Training and Aiding. Final Technical Paper for Period February-December 1989.
INSTITUTION	Search Technology, Inc., Norcross, GA.
SPONS AGENCY	Air Force Human Resources Lab., Brooks AFB, TX. Training Systems Div.
REPORT NO	AFHRL-TP-89-78
PUB DATE	May 90
CONTRACT	F33615-86-C-0545
NOTE	66p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Artificial Intelligence; *Cost Effectiveness; Decision Making; Evaluation Methods; *Helping Relationship; Learning; Military Training; Models; *Predictive Measurement; Problem Solving; Program Design; Research Methodology; *Training
IDENTIFIERS	*Tradeoffs

ABSTRACT

A methodological framework is presented for representing tradeoffs among alternative combinations of training and aiding for personnel in complex situations. In general, more highly trained people need less aid, and those with less training need more aid. Balancing training and aiding to accomplish the objectives of the system in a cost effective way is the concern. A wide variety of methods, tools, and models is reviewed. These approaches are evaluated in terms of their advantages and disadvantages when used to analyze training/aiding tradeoffs. The use of the proposed framework and its component methods, tools, and models is illustrated by an analysis of a realistically complex example involving the design of a head-up display for use by truck drivers in long-haul transport. Results demonstrate that the tradeoff issue can be involved in other than an ad hoc manner. Research needed in predictive models, learning processes, and intelligent systems is reviewed. Four tables and nine figures illustrate the discussion. (SLD)

TM

AFHRL-TP-89-78

AIR FORCE HUMAN RESOURCES LABORATORY

COMPUTATIONAL APPROACHES FOR ANALYZING TRADEOFFS BETWEEN TRAINING AND AIDING

William B. Rouse
William B. Johnson

Search Technology, Incorporated
4725 Peachtree Corners Circle
Norcross, Georgia 30092

TRAINING SYSTEMS DIVISION
Brooks Air Force Base, Texas 78235-5601

May 1990
Final Technical Paper for Period February – December 1989

Approved for public release; distribution is unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

TM 020216

NOTICE

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

THEODORE A. LAMB
Contract Monitor

HENDRICK W. RUCK, Technical Advisor
Training Systems Division

RODGER D. BALLENTINE, Colonel, USAF
Chief, Training Systems Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
May 1990

3. REPORT TYPE AND DATES COVERED
Final - February - December 1989

4. TITLE AND SUBTITLE
Computational Approaches for Analyzing
Tradeoffs Between Training and Aiding

5. FUNDING NUMBERS
C - F33615-86-C-0545
PE - 62205F
PR - 1121
TA - 11
WU - 00

6. AUTHOR(S)
William B. Rouse
William B. Johnson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Search Technology, Incorporated
4725 Peachtree Corners Circle
Norcross, Georgia 30092

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Training Systems Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235-5601

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFHRL-TP-89-78

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This paper presents a methodological framework for representing tradeoffs among alternative combinations of training and aiding for personnel in complex systems. A wide variety of methods, tools, and models are reviewed. These approaches are evaluated in terms of their advantages and disadvantages when used to analyze training/aiding tradeoffs. These evaluations lead to the synthesis of three composite approaches to analyzing tradeoffs. The use of the proposed framework and its component methods, tools, and models is illustrated by analysis of a realistically complex example.

14. SUBJECT TERMS
aiding
modeling
training

15. NUMBER OF PAGES
66

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std Z39-18

AFHRL Technical Paper 89-78 May 1990

COMPUTATIONAL APPROACHES FOR ANALYZING TRADEOFFS BETWEEN TRAINING AND AIDING

William B. Rouse
William B. Johnson

Search Technology, Incorporated
4725 Peachtree Corners Circle
Norcross, Georgia 30092

TRAINING SYSTEMS DIVISION
Brooks Air Force Base, Texas 78235-5601

Reviewed and submitted for publication by

James B. Bushman, Major, USAF
Chief, Training Assessment Branch

This publication is primarily a working paper. It is published solely to document work performed.

SUMMARY

This paper presents a methodological framework for representing tradeoffs among alternative combinations of training and aiding for personnel in complex systems. A wide variety of methods, tools, and models are reviewed. These approaches are evaluated in terms of their advantages and disadvantages when used to analyze training/aiding tradeoffs. These evaluations lead to the synthesis of three composite approaches to analyzing tradeoffs. The use of the proposed framework and its component methods, tools, and models is illustrated by analysis of a realistically complex example.

C

PREFACE

This paper is concerned with analyzing tradeoffs in the design of complex systems. Various aspects of systems are designed to aid humans in those systems as they perform their tasks. Examples of aids are manuals, test equipment, and enhanced cockpit displays. Another important factor in the design of systems is the training humans receive to perform their tasks.

In general, more highly-trained people need less aiding, and those with less training require more aiding. The research question is how should one balance training and aiding to accomplish the operational objectives in a cost effective manner? The goal of this paper is to contribute to answering this question by focusing on computational approaches to analyzing the tradeoffs between training and aiding.

Future research will focus on identifying the additional characteristics of tasks which contribute to deciding whether tasks should be job aided, trained, or some combination of the two and how the training or job aiding should be accomplished, that is, what does the job aid do or what type of training is best.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge many insightful and helpful comments and suggestions from Drs. James Bushman, Theodore Lamb, Scott Newcomb, and Hendrick Ruck. The preparation of this paper was supported by the Air Force Human Resources Laboratory under Contract No. F33615-86-C-0545.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

This paper is concerned with analyzing a particular class of tradeoffs in the design of complex systems. The term design is used here to denote the broad set of activities associated with conceptualizing and developing a system such as an aircraft or power plant, as well as the planning associated with staffing, training, and supporting the personnel involved with the system.

One specific tradeoff is of particular interest in this paper. Various aspects of systems are designed to aid humans in those systems as they perform their tasks. Examples include manuals, test equipment, and enhanced cockpit displays. A related aspect of design concerns training humans to perform their tasks.

In general, more highly-trained people need less aiding, and those with less training require more aiding. The tradeoff is obvious. How should one balance training and aiding to accomplish the operational objectives of the system in a cost effective manner? The goal of this paper is to contribute to answering this question.

To the extent that this tradeoff has been explicitly addressed in the past, analysis has relied heavily on past experiences with similar systems. Typically, these types of analysis have proceeded once detailed design has been completed. Further, they usually have required many person-years of effort. Often, the result has been a time-consuming and expensive effort that provided insights which were too late to be implemented in any substantial way.

This paper focuses on computational approaches to analyzing tradeoffs between training and aiding. We envision one or more analysts, perhaps at individual workstations or, alternatively, linked via networked workstations, accessing and utilizing a variety of computational methods and tools for the purpose of identifying, structuring, and analyzing training/aiding tradeoffs. This paper rationalizes and develops this vision, as well as illustrates its application to a realistically complex example.

The Problem

It is useful to crisply summarize the problem being addressed. The tradeoff of interest concerns the relative emphasis on creating the potential to perform (via training) versus augmenting performance directly (via aiding). More simply, there is a tradeoff between putting "smarts" in people versus putting "smarts" in machines. This tradeoff is becoming increasingly central as progress in the area of machine intelligence continues to evolve.

The general requirements for a computational approach to resolving this tradeoff include computer-based methods and tools for:

1. Predicting the impact of training/aiding alternatives on human performance, as well as system and mission performance.
2. Including personnel-related considerations (e.g., aptitude requirements) in tradeoff analyses.
3. Including manpower-related considerations (e.g., staffing requirements) in tradeoff analyses.

This paper explores in detail alternative methods and tools in these areas.

Decision Making Context

It is essential that we recognize the broader context within which any realistic training/aiding tradeoffs must occur (Akman Associates, 1987; Booz-Allen & Hamilton, 1985; Thurman, 1989). Figure 1 depicts several relationships among design, training, and staffing.

In the left column, design proceeds from requirements to performance, typically with the assumption that human behavior will satisfy task requirements. The training process (center column) is responsible for providing people who can meet these expectations. The manpower and personnel process (right column) is responsible for producing a sufficient number of trainable people to satisfy mission requirements.

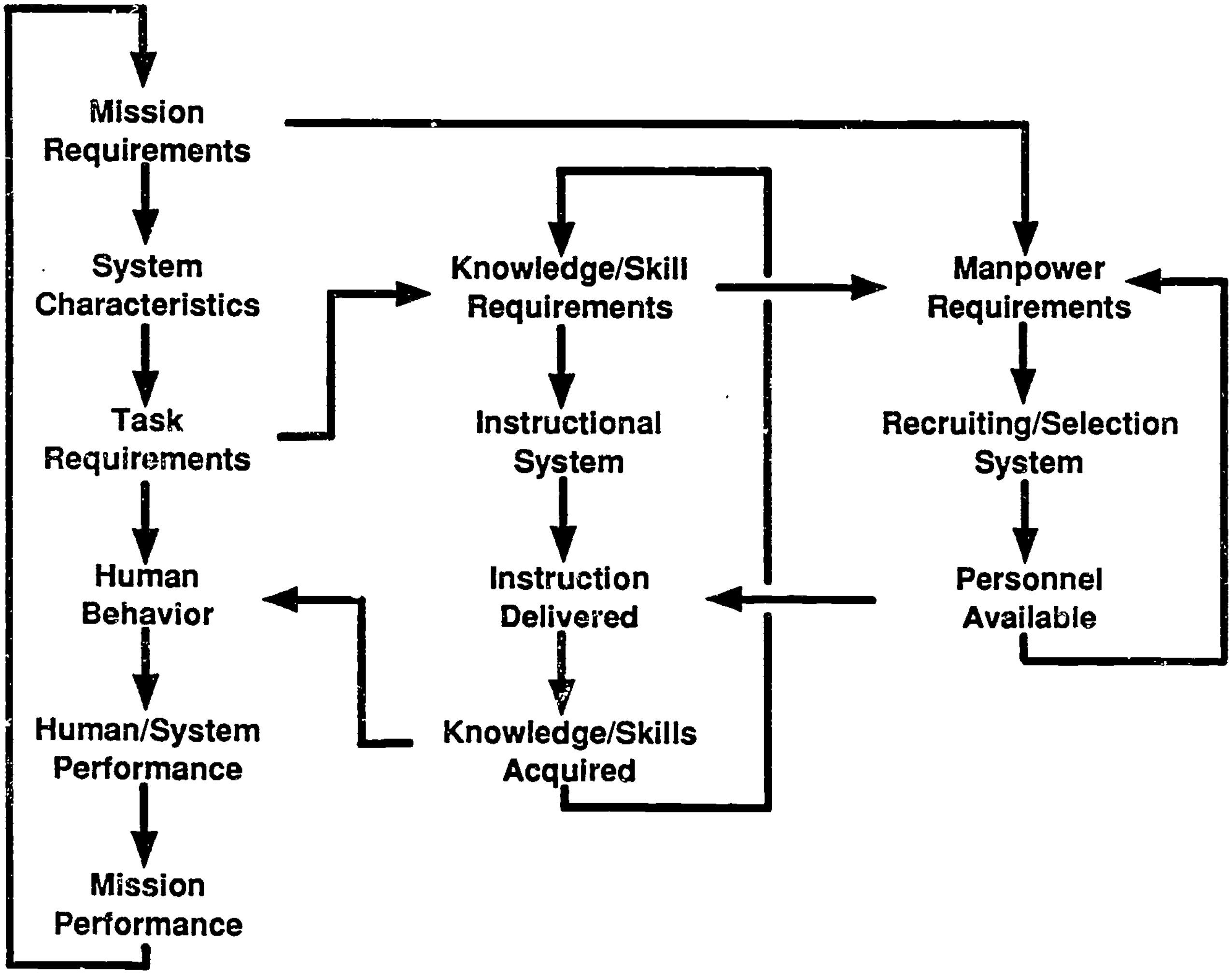

Figure 1. Design, Training, and Staffing.

One might expect there to be a rich interplay among the three processes depicted in Figure 1. For example, designers might ask about the "trainability" of people for task requirements emerging from new technologies embedded in system characteristics. Similarly, training specialists might ask about the "recruitability" of people with particular aptitudes. The answer to the recruitability question might greatly influence the answer to the trainability question.

This interplay of design, training, and staffing might fit into an overall process such as shown in Figure 2. Decision making at the highest level in this process is concerned with the tradeoff between mission effectiveness and life-cycle costs, as they relate to mission requirements. Lower-level tradeoffs and decision making cascade to produce the higher-level measures of interest. For example, manpower, personnel, and training (MPT) parameters and their resulting costs are "driven" by earlier decisions.

Ideally, downstream impacts would be fed back upstream to modify, for example, design decisions that have highly undesirable MPT impacts. In practice, this seldom occurs because of the temporal and organizational separation of these issues. Temporal separation occurs because downstream decisions are often not pursued until upstream decisions are made. The problem is that upstream decision makers have few predictive methods and tools to enable them to project the downstream impacts of their decisions while they are still reversible or modifiable. This paper is concerned with providing such methods and tools - particularly for the cross-hatched elements of Figure 2.

The organizational separation of issues results in "suboptimization" in the sense that each issue and associated tradeoffs are resolved in a locally optimal manner, which may undermine the possibility of global optimization. While some of this organizational separation is due to historical precedents and political expedients, it seems reasonable to assert that much of the organizational decomposition (and hence separation) reflects an attempt to cope with the complexity of designing large-scale, advanced systems.

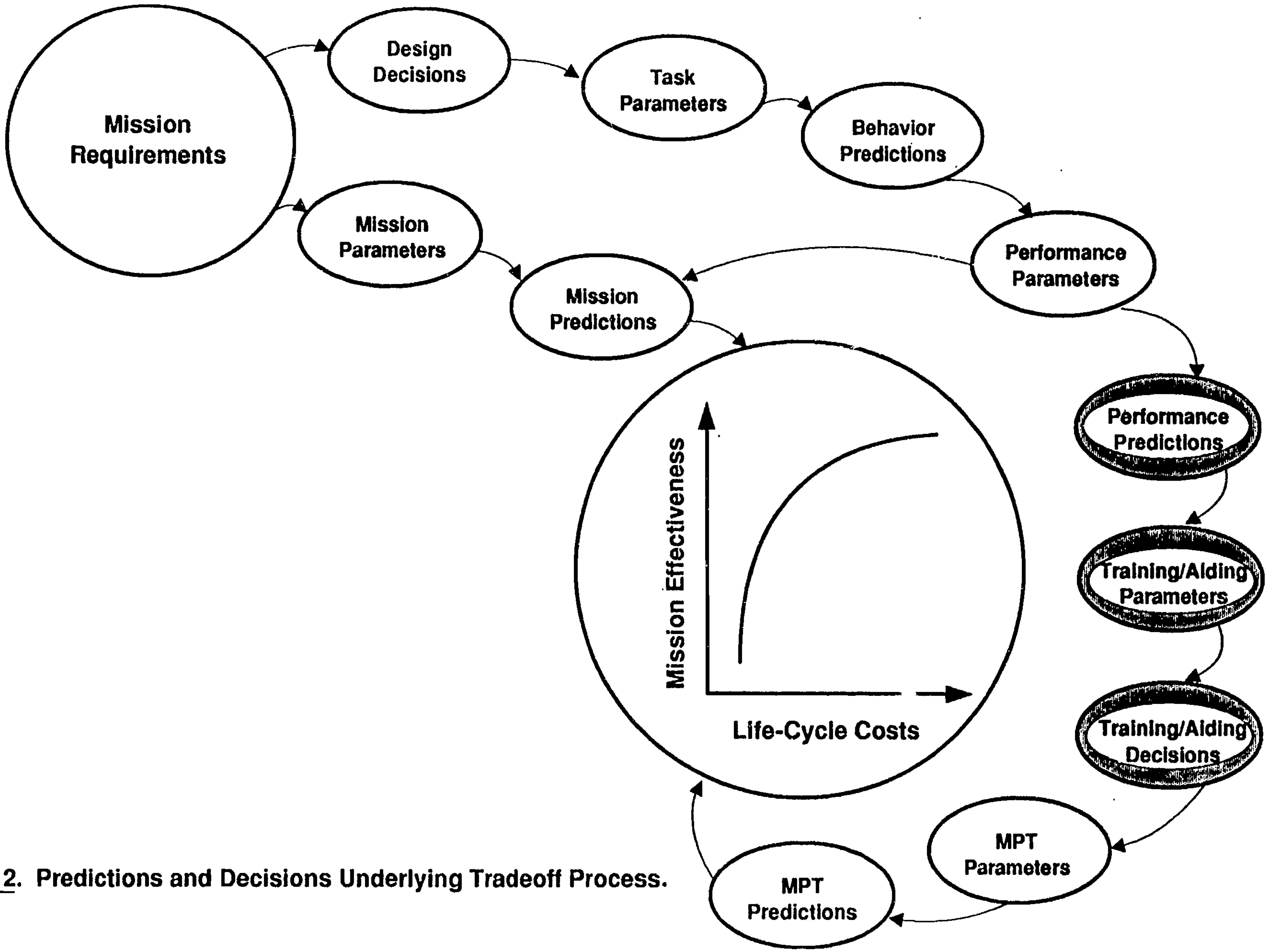

Figure 2. Predictions and Decisions Underlying Tradeoff Process.

Appropriate methods and tools may be able, in effect, to decrease this complexity and enable consideration of more global tradeoffs. The approach discussed in this paper represents an effort in this direction.

It is important to emphasize the fact that the approach presented here builds on a rich history of research on training and aiding. There are many ways to train and aid people, and many data available evidencing the benefits of the alternatives. What is lacking, however, is a way of trading off alternatives without actually performing an empirical study. This difficulty is due to the absence of a framework for integrating disparate results and performing tradeoff analyses. The key to developing such a framework is incorporating means for predicting, rather than measuring, the impact of selecting particular training and aiding alternatives. This paper presents an approach to achieving this goal.

An Example

The methodological framework presented in this paper is fairly comprehensive and somewhat abstract. As such, there is a risk that many readers will perceive this framework to be applicable to everything in general and nothing in particular. To avoid this possibility, we have developed a realistically complex example that is elaborated throughout this paper.

At this point, we will limit the discussion to introducing the context of the example. The problem of interest concerns the design of a head-up display (HUD) for use by truck drivers in long-haul transport operations. This problem emerged from a fictitious client's desire to reduce truck downtime due to weather and rerouting for maintenance.

Discussions with the client led to the definition of three overall objectives. The primary objective was to enable all-weather operations--in other words, continued high-speed operations in rain, fog, and snow. Second, whatever equipment was added to the truck to achieve this capability should be sufficiently reliable and maintainable to not require any rerouting for specialized maintenance. Finally, the client wanted a "high tech" solution that could be advertised as part of the company's long-standing image as an industry innovator.

The client's small engineering staff had studied alternative ways to meet these objectives and concluded that they needed external expertise to proceed. In the process of their studies, they had seen head-up displays that are now available as options on a few new automobiles, as well as a variety of airplanes. The primary value of a HUD is that it projects displayed information on the inside surface of the windshield and, therefore, the driver can see this information without taking his or her eyes off the road.

The client's engineers thought that it might be possible to use a HUD to provide information that would enable a driver to stay on the road, avoid obstacles, and maintain speed, even though rain, fog, or snow had substantially reduced visibility. The client wanted us to develop and evaluate this concept to determine if it could accomplish the aforementioned three objectives.

Further elaboration of the HUD concept and its implications is deferred until a later section of this paper. However, it is important to note several training and aiding considerations that emerge in this later discussion. First, the HUD is basically an aid for truck operations. Various levels of sophistication of aiding are possible. Each of these levels has implications for the training required to use the HUD successfully. Two clear tradeoffs emerge in the analysis of operations using the HUD.

There are also maintenance tasks associated with the HUD concept. In order to realize the full potential of the concept, as well as avoid rerouting for specialized maintenance, the truck driver will have to be able to perform some level of maintenance. This obviously has training implications. Further, it is possible to provide aiding for some aspects of the maintenance tasks. As might be expected, aiding possibilities and training requirements interact. The result for this example is two additional tradeoffs that are elaborated in later discussions.

II. ALTERNATIVE APPROACHES

There are at least three distinctly different ways that the resolution of training/aiding tradeoffs can be approached. Similarly, there are three types of information source that can support each of these approaches. In this section, the three approaches and three information sources are discussed. Later in this paper, hybrid or composites of these approaches/sources are considered and evaluated.

One approach to training/aiding tradeoffs involves compiling general guidelines for training/aiding decisions based on cumulative experience and experiments. Such guidelines map the attributes of a training/aiding situation to combinations of specific types of training and aiding. With this approach, tradeoffs are implicit in the guidance. However, the user of the guidelines does not explicitly formulate tradeoffs. Thus, to a great extent, decision making is proceduralized, e.g., if situation x, then employ training type y and aiding type z.

The second alternative to resolving training/aiding tradeoffs involves predicting human and system performance as a function of training/aiding alternatives, and using these predictions as a basis for tradeoffs. This approach requires an analyst to explicitly formulate tradeoffs in terms of independent and dependent variables - in other words, characteristics of alternatives that can be manipulated and measures of the impact of these manipulations. Further, this approach requires explicit comparison of performance predictions across alternatives, and explicit assignment of relative benefits and costs to these predictions.

The third approach to resolving tradeoffs involves simulating human and system behaviors as a function of training/aiding alternatives, calculating performance measures based on these behaviors, and using these calculations as a basis of tradeoffs. Once performance measures are calculated, the process of resolving tradeoffs proceeds in a manner similar to that based on performance predictions. The essential difference with using behavioral simulation is the emulation of the process of actually receiving training or using aiding - in contrast, performance predictions are only concerned with the results of this process.

Information Sources

In order to predict behavior or performance, as well as follow guidelines, a variety of information is needed. There are three sources of this information: judgment, archives, and models. Judgment includes the opinions, preferences, and observations of an analyst himself or herself, colleagues, and subject-matter experts (SMEs). Judgment is the most frequently used source of information in many technical domains (Allen, 1977; Rouse & Cody, 1988). The reason is quite simple - judgment is readily accessible, easily consumable, and provides answers that are "good enough."

Archives include data bases, fact sheets, handbooks, text books, and journals. Archives usually include the types of information that researchers generate, compile, and communicate. Typically, practitioners find this information difficult to access (e.g., from a library), requires much effort to consume (i.e.,study), and provides answers that are correct in general but may be a bit off-target in particular.

Models are means for generating information by approximating the characteristics and processes underlying the application of interest. Types of model include: experiential, empirical, and analytical. An experiential model of a new system might be the previous version of that system (i.e., a baseline) and a characterization of how the new system will be different. An experiential model provides answers to "what if" questions by assuming that the new system's behavior and performance will be very similar to that of the old system, except for the upgrades envisioned to overcome past deficiencies or provide new capabilities.

An empirical modeling effort involves collecting data for conditions and subject populations that are assumed to represent the eventual operating conditions and user population of the system o interest. The behaviors observed and performance measures calculated are useful to the extent that experimental conditions and subject populations are good models of the target application. For this reason, empirical data are not inherently more accurate than, for example, expert judgments of likely behaviors or performance.

Analytical modeling involves constructing a computational representation of the processes of interest and computing various characteristics of this representation, typically its response to various manipulations. The processes represented may vary in levels of abstraction and aggregation. For example, a representation might model basic psychological processes such as memory and reaction time, or more aggregate phenomena such as learning curves. As another illustration, a model might represent fairly concrete human/system behaviors such as manual control, or more abstract phenomena such as mission effectiveness.

Approaches versus Sources

Summarizing briefly, we have discussed three general approaches to resolving training/aiding tradeoffs (i.e., guidelines, performance predictions, and behavioral simulations). We also have discussed three general sources of information for applying these approaches (i.e., judgment, archives, and models). Table 1 illustrates how the three approaches and three sources combine to provide alternative methods, tools, and models for addressing training/aiding tradeoffs.

This tabulation is reasonably self-explanatory. It is useful, however, to point out a few general characteristics of these alternatives. Judgment tends to be used to produce qualitative outputs, even if the inputs to the process are quantitative. This is often exactly what is needed for many types of decision. Archival information is usually more

Table 1. Alternative Methods, Tools, and Models

	BEHAVIOR PREDICTIONS	PERFORMANCE PREDICTIONS	TRAINING/AIDING GUIDELINES
● JUDGMENT - Self - Colleagues - SMEs	Qualitative predictions of categories of behavior via past experiences	Qualitative predictions of relative perf. via past experiences	Structured mappings of human/system characteristics to T/A decisions
● ARCHIVES - Data bases - Fact sheets - Handbooks - Textbooks - Journals	Qualitative predictions of categories of behavior via psych. theories	Quantitative predictions via handbooks and other compilations	Structured mappings of expt. results to training/ aiding decisions
● MODELS - Experiential - Empirical - Analytical	Quantitative predictions via psychological models and/or expt. paradigms	Quantitative predictions via MMS models, expt. studies, and/or baselines	Structured mappings of predictions to training/aiding decision process

quantitatively presented, but in a relatively context-free manner. The analyst, therefore, must qualitatively "adjust" the information for the particular problem - the apparent precision cannot actually be used. Models are often quantitative, in terms of both inputs and outputs, and usually are tailored to particular contexts. Nevertheless, an inappropriate representation usually is not sufficiently adjustable to compensate for a poor choice.

Tables 2 and 3 show specific examples of the general methods, tools, and models shown in Table 1. (Note that Table 3 is an expansion of the lower left of Table 2 which is outlined in bold) The examples shown in Tables 2 and 3 were chosen to be representative - they are, in our opinion, the best exemplars of the approaches they embody. Many other methods, tools, and models could be cited; however, such additions would not significantly broaden the range of approaches represented.[1]

Discussion of each specific entry in Tables 2 and 3 is delayed until a later section on evaluation of alternatives. However, it is important to note the large number of alternatives available. Our "toolbox" is very full. We now need a means of matching problems to methods, tools, and models.

III. SELECTION CRITERIA

It is easy to imagine a variety of criteria that might influence the selection of a particular method, tool, or model. To determine what criteria users actually employ in this selection process, we assessed the preferences of participants at a NATO Workshop on Applications of Human Performance Models to System Design, held in Orlando, Florida, in April 1988. The workshop included presentations of 29 types of method, tool, or model - the papers upon which these presentations were based appear in McMillan (1989).

---

[1] Discussions of a wide range of methods, tools, and models can be found in Baron and Kruser (1988); Elkind, Card, Hochberg, and Huey (1989); Fleger, Permenter, and Malone (1987); McMillan (1989); Moraal and Kraiss (1981); Rouse (1980); and Sheridan and Ferrell (1974).

Table 2. Example Methods, Tools, and Models

	BEHAVIOR PREDICTIONS	PERFORMANCE PREDICTIONS	TRAINING/AIDING GUIDELINES
JUDGMENT	• Prototyping and demonstration methods (Wasserman & Shewmake, 1985)	• Cognitive Requirements Model (Rossmeissl et al., 1989)	• Task profile ratings (Irvin et al., 1988)
ARCHIVES	• Human Performance Handbook (Boff et al., 1986) • Learning theory (Glaser & Bassok, 1989) • Problem solving theory (Greeno & Simon, 1988)	• Human Performance Compendium (Boff & Lincoln, 1988) • Meta-analyses for computer-based instruction (Kulik & Kulik, 1988) • Troubleshooting review (Morris & Rouse, 1985)	• Job performance aid selection algorithm (Booher, 1978) • Training/aiding decision flowchart (Foley, 1978) • Integrated Personnel System development model (Smillie & Blanchard, 1986)
MODELS	• Experimental studies	• Baseline systems • Experimental studies	• Comparability analysis - Hardman (Weddle, 1986) • Resource allocation model (Rouse, 1985) • Integrated support system tradeoff model (Rouse, 1987)
	• Operator models	• Operator models	
	• Maintainer models	• Maintainer models	

Table 3. Example Operations and Maintenance Models

	BEHAVIOR	PERFORMANCE
BOTH	• Experimental studies	• Baseline systems • Experimental studies
OPERATIONS	• Ladder model, etc. (Rasmussen, 1986) • KARL model (Knaeuper & Rouse, 1985) • PROCRU model (Baron, 1984)	• HOS model (Lane et al., 1981) • Manual control models (McRuer et al., 1965; Kleinman et al., 1971) • Learning curve models (Towill, 1989)
MAINTENANCE	• Ladder model, etc. (Rasmussen, 1986) • Fuzzy rule-based model (Hunt & Rouse, 1984) • Profile model (Towne et al., 1982	• Maintainability models (Goldman & Slattery, 1964) • Complexity models (Rouse & Rouse, 1979; Wohl,1982) • Petri net model (Madni et al., 1984) • MAPPS model (Siegel et al., 1984) • Learning curve models (Towill, 1989)

There were 140 participants. Participants received a structured questionnaire, which asked them to rate each type of method, tool, or model in terms of seven criteria. They were also asked to indicate their likely subsequent behavior toward the method, etc. Behaviors among which they could choose included seeking more information, advocating use in their organization, or intending use themselves. They could, of course, also indicate no interest. Approximately 100 questionnaires were returned.

The rich set of data resulting was analyzed in a variety of ways, and yielde , many insights into what does or does not influence potential users' opinions (Cody & Rouse, 1989; Rouse & Cody, 1989a). Of particular interest here are the criteria that most influence users' preferences. The questionnaires indicated that users' felt that six of the seven criteria strongly influenced their decision making - only cost was given a low weighting. However, discriminant analyses to determine the weightings of criteria that best discriminated likely subsequent behaviors showed that only two criteria accounted for roughly 80% of the variance in users' expected behaviors. These two criteria were applicability and availability. In other words, users were primarily interested in the extent to which a method, etc. was relevant to their problems, and the extent to which they could readily access it.

Table 4 indicates the general nature of applicability and availability for the types of information source compiled in Tables 1, 2, and 3. The nature of availability is quite straightforward and does not differ substantially among the types of source. In contrast, there are very significant differences among the ways applicability should be assessed for the three types of information source.

Evaluation Questions

These differences are best illustrated by the specific questions compiled in Figures 3 through 6. Questions related to evaluating judgment as an information source are shown in Figure 3. If one is considering using the methods and tools which employ

Table 4. Applicability and Availability of Information Sources

INFO SOURCE	APPLICABILITY	AVAILABILITY
● JUDGMENT - Self - Colleagues - SMEs	● Relevant experience ● Acceptable accuracy ● Acceptable resolution	● Avail. in-house ● Avail. via telephone ● Approachable
● ARCHIVES - Data bases - Fact sheets - Handbooks - Text books - Journals	● Relevant conditions ● Appropriate population ● Appropriate measures ● Acceptable accuracy ● Acceptable resolution	● Avail. in-house ● Avail. via databases
● MODELS - Experiential - Empirical - Analytical	● Phenomena representable ● Appropriate outputs ● Available inputs ● Available accuracy ● Acceptable resolution	● Avail. in-house ● Avail. packages

Applicability

- Are the experiences of the SMEs relevant?
- Are the levels of expertise acceptable?
- Are any biases apparent?
- Are the impacts of these biases acceptable?
- Is the accuracy of judgments acceptable?
- Is the repeatability/reliability of judgments acceptable?
- Is the resolution of judgments acceptable?
- Will judgments fairly consider new/novel concepts?

Availability

- Are SMEs accessible in-house or via telephone?
- Are they approachable?
- Are they able to express their opinions?

Figure 3. Evaluating Judgment as an Information Source.

Applicability

- Are the conditions of data collection relevant?
- Are the ways in which theories apply apparent?
- Are the populations for which data were collected (or theories developed) appropriate?
- Are the measures employed appropriate?
- Will the level of aggregation employed allow sufficiently accurate estimates for conditions of interest?
- Is the power of tests of null results sufficient?
- Is the level of quantification of guidelines commensurate with decisions of concern?
- Is the impact of <u>not</u> following guidelines quantified at a level commensurate with decisions of concern?

Availability

- Are compilations accessible in-house or via data bases?
- Are these compilations easily accessible?
- Are these compilations understandable in terms of concepts, jargon, and perspectives?
- Are data for null results available?

Figure 4. Evaluating Archives as an Information Source.

Applicability

- Are there appropriate baseline systems? (e.g., or is the new system/technology novel?)
- Are there appropriate experimental methods? (e.g., or is data collection not viable?)
- Are analytical formulations compatible with phenomena to be represented? (e.g., can the process of learning be represented?)
- Are the outputs of models appropriate for the decisions of concern? (e.g., or are they too aggregated or use the wrong metrics?)
- Are the data for the baseline system available?
- Are appropriate experimental conditions known?
- Are parameter estimates feasible for analytical models?
- Are sensitivity analyses reasonable substitutes for inadequate data?
- Is the accuracy of the models' outputs acceptable?
- Is the resolution of the models' outputs acceptable?

Availability

- Is information about the baseline system accessible in-house or elsewhere?
- Are experimental facilities accessible in-house or elsewhere?
- Are software packages for analytical models accessible in-house or elsewhere?
- Is expertise on the baseline system, use of facilities, or use of modeling packages accessible in-house or elsewhere?

Figure 5. Evaluating Models as an Information Source.

Judgment

- Is the accuracy of judgments assessable?
- Is the repeatability/reliability of judgments assessable?
- Is the resolution of judgments assessable?

Archives

- Are compilations understandable in terms of concepts, jargon, and perspectives?
- Is the impact of <u>not</u> following guidelines included?
- Are data for null results available?
- Is the power of tests of null results sufficient?

Models

- Are analytical formulations compatible with phenomena to be represented? (e.g., can the process of learning be represented?)
- Are parameter estimates feasible for analytical models?
- Are sensitivity analyses reasonable substitutes for inadequate data?
- Is the accuracy of the models' outputs assessable?
- Is the resolution of the models' outputs assessable?
- Are software packages for analytical models accessible in-house or elsewhere?

<u>Figure 6</u>. **Questions Whose Answers Are Not Totally Application Dependent.**

judgment as an information source, (i.e., those in the top row of Table 2), then the questions in Figure 3 are particularly important. These questions are not easy to answer, and we suspect are seldom asked. Specific answers to these questions are considered in a later section where the alternatives in Tables 2 and 3 are evaluated.

Figure 4 shows questions related to evaluating the archives as an information source. These questions are quite different than those in Figure 3, reflecting the differences between judgment and archives as information sources (i.e., between the top and center rows of Table 2). As noted earlier, a primary difficulty with archival information sources is that conditions and subject populations for which data were collected may not match the problem at hand. Thus, the accuracy for the situation of interest may be doubtful, despite the apparent high resolution of the compilations presented.

Another difficulty with archival information is the typical lack of conclusions regarding what variables do <u>not</u> affect the measures of interest (i.e., null results). In a recent study of designers' information needs (Rouse & Cody, 1988), participants frequently noted a need to determine which of a large number of variables were unimportant, so that they could focus their empirical efforts on critical variables. Unfortunately, the archives seldom contain <u>defendable</u> conclusions concerning null results. This issue and others related to Figure 4 are further considered in later discussions of evaluating specific alternatives.

Questions concerning evaluation of models as information sources are shown in Figure 5. These questions relate to evaluating the entries in the bottom row of Table 2 and all of Table 3. Some of the questions related to experiential and empirical models are analogous to several of the questions concerning archival compilations of data. The questions for analytical models are quite different.

The best illustration of this difference concerns the extent to which the type of representation underlying a particular analytical formulation (e.g., networks or differential equations) is compatible with the specific phenomena to be represented. This issue is particularly important if one is concerned with training. Many of the models in Tables 2 and 3 cannot reasonably be used to represent the process of learning (rather that the product). To a great extent, this is due to our limited knowledge of this process. However, this limitation is also related to the nature of many of the types of representation underlying analytical models which, at the very least, are extremely cumbersome when attempting to emulate learning. While this limitation is important, the situation is not totally bleak, as is illustrated in later discussions of specific models.

Another issue related to analytical models concerns estimating the parameters within the structure of a model, including obtaining data upon which these estimates can be based. This can present difficulties when parameters are "internal" to a phenomenon in the sense that they inherently are not measurable and must be inferred, perhaps via some least-squares method. This limitation is likely to be acceptable if one is only concerned with the input/output predictive validity of the model. However, if one's goal is to make inferences about the process underlying the input/output, then the possibility of having non-unique parameter estimates can preclude any strong assertions about construct validity (Rouse, Hammer & Lewis, 1989). Fortunately, predictive validity is likely to be sufficient for the types of training/aiding tradeoffs addressed in this paper.

Most of the questions in Figures 3, 4, and 5 cannot be answered without consideration of the particular way in which a method, tool, or model is to be used. Constraining the "application space" to analyzing training/aiding tradeoffs provides some focus, but not enough to provide context-free answers. Figure 6 shows the subset of questions that can be addressed in some meaningful way without further constraining the application space. In the next section, this subset is used to evaluate the methods, tools, and models in Tables 2 and 3.

IV. EVALUATION OF ALTERNATIVES

This section is organized as follows. Evaluation results are presented according to the rows of Table 2, using successive subsections for judgment, archives, and models. Within subsections, discussions proceed from left to right in Table 2. Within each of the nine (i.e., 3x3) sections of Table 2, each list is discussed from top to bottom. Note that Table 3 is an expansion of the two leftmost elements of the row labeled Models in Table 2. Thus, entries in Table 3 are discussed in the order dictated by Table 2.

The remainder of this section presents the results of applying the questions in Figure 6 to the entries in Tables 2 and 3 in the aforementioned order. The general nature of our evaluative comments focuses on the strengths and weaknesses of each approach, where the entries in Tables 2 and 3 are taken to be exemplars of the approach. It is not possible within the scope of this paper to provide other than a cursory description of each of the approaches in Tables 2 and 3. The later discussion of the example provides more detail on a subset of these approaches. However, more comprehensive and detailed treatments must necessarily be sought from the documents in the reference list.

Judgment

Prototyping and Demonstration Methods (Wasserman & Shewmake, 1985)

While users' reactions can be of great value, it often is not clear if available "users" are representative of eventual users. For this reason, users may be of most value for explaining their current tasks and environment, as well as evaluating approaches to supporting their current ways of performing their tasks (Rouse & Cody, 1989b). Users are likely to be of less value for evaluating new ways to perform tasks. Finally, users are likely to offer apparent accuracy and resolution in excess of what can be objectively justified.

Cognitive Requirements Model (Rossmeissl et al., 1989)

The structured nature of this approach makes it more likely that the results of using it will be repeatable. Further, previous evaluation of the approach increases confidence in its use. The predictions of cognitive requirements are qualitative involving, for example, ordinal comparisons of alternatives, or comparison of an alternative to a baseline. Regardless of the rigor of the assessment procedure or the scales chosen, quantitative predictions are difficult due to the accuracy and resolution limitations of the judgments that serve as inputs.

Task Profile Ratings (Irvin et al., 1988)

This approach focuses on producing training/aiding recommendations rather than projecting quantitative performance implications of task characteristics. Accuracy and resolution are lesser issues since the mapping is to a few training/aiding alternatives. However, the coarseness of this mapping limits the possibility of fine-grained design tradeoffs, e.g., simulator features versus training transfer. Thus, this approach to guidelines is likely to be of most use in situations where perpetuation of previous training/aiding concepts is desirable and appropriate.

Archives

Human Performance Handbook (Boff et al., 1986)

The contents of this research-oriented handbook seem very relevant in general, but it is not always clear how to apply this information to specific problems. There is a reasonable level of quantification. However, uncertainty about relevance of conditions and subject populations may undermine the value of the accuracy and resolution of the data. Null results are occasionally noted, at least qualitatively. Users of this book must be relatively sophisticated in terms of understanding behavioral science concepts and methods.

Learning Theory (Glaser & Bassok, 1989)

Seems to be very relevant to understanding the learning processes underlying training. Unfortunately, the state of the art is such that it is not clear how this conceptual and qualitative information can be applied to specific problems. Users of this material must be very sophisticated relative to behavioral science concepts and methods.

Problem Solving Theory (Greeno & Simon, 1988)

Seems very relevant to aspects of aiding, but it often is not clear how to apply this conceptual and qualitative information to realistically complex problem solving situations. Requires very sophisticated users.

Human Performance Compendium (Boff & Lincoln, 1988)

Much more easily usable version of its "cousin," the *Human Performance Handbook*. The presentation is generally crisper and more quantitative than the handbook. While the contents remain relatively context free, the focus of the *Compendium* on engineering usage partially compensates for this. Null results are occasionally presented and/or discussed. The *Compendium* is reasonably user-oriented relative to the engineering designer.

Meta-Analyses for Computer-Based Instruction (Kulik & Kulik, 1988)

Much more focused than the *Compendium*. Accuracy and levels of resolution are reasonable for interpolation among the results of many studies of computer-based instruction. Extrapolation outside the range of these studies is questionable. Contents are relatively context free, but focus on computer-based instruction compensates partially. There is some discussion of null results.

Troubleshooting Review (Morris & Rouse, 1985)

This review is an exemplar of many state-of-the-art reviews. While the results discussed cross many studies, the procedure for integrating results and drawing conclusions is not as rigorous as meta-analysis. As a result, the conclusions are fairly qualitative. Discussion of null results is limited to statements such as "there was insufficient evidence to reject the null hypothesis." Because the topic is very focused, the issue of context is less a concern.

Job Performance Aid Selection Algorithm (Booher, 1978)

Uses qualitative characteristics of people, tasks, and systems to provide qualitative guidance for training/aiding decisions. Guidelines are based on performance results, but do not support explicit performance tradeoffs. Resolution is very low, for the most part due to the intent of the guidelines. The implications of not following the guidelines are unclear. Guidelines are easily understandable, although assessment of input variables may be difficult. Guidelines implicitly inform of null results by not including them as input variables. This approach is of most use in situations where perpetuation of previous training/aiding concepts is desirable and appropriate.

Training/Aiding Decision Flowchart (Foley, 1978)

This approach is very similar, in objectives and format, to the job performance aid selection algorithm. Consequently, the answers to the evaluation questions are basically the same.

Integrated Personnel System Development Model (Smillie & Blanchard, 1986)

Prescribes a particular approach for integrating training and aiding that is conceptually quite broad. This approach integrates past data and experiences to outline a comprehensive support concept, rather than a set of decision rules for choosing among training/aiding alternatives. In this manner, this approach provides guidance on how alternatives should fit together, as opposed to how specific training/aiding elements should be chosen.

Models

Experimental Studies

If a key study is identified, then a simulator study or even a field study can be invaluable in terms of accuracy, resolution, credibility, etc. Otherwise, such studies are too expensive and too slow. It is also sometimes difficult to assure representative conditions and populations.

Ladder Model and Related Constructs (Rasmussen, 1986)

This qualitative model is relevant to both operations and maintenance tasks. Accuracy and resolution for this model relate to predictions of categories of behavior, rather than specific sequences of behaviors. This model can provide a framework for mapping to quantitative models. Otherwise, predictions tend to be qualitative and/or weak.

KARL Model (Knaeuper & Rouse, 1985)

KARL (Knowledgeable Application of Rule-Based Logic) is a rule-based model of operator behavior that can provide good predictions of sequences of behavior if an appropriate knowledge base is encoded. Aiding and training are natural addenda - however, this has thus far been limited to procedural aiding. While learning can be represented in terms of rule acquisition, it is not clear how this construct would be validated. Resolution is not an issue; accuracy is difficult to project. Knowledge acquisition can be problematic; parameter estimation is usually not central. There is relatively little experience upon which to base decisions about simulation issues such as number of runs, appropriate statistics, etc.

PROCRU Model (Baron, 1984)

PROCRU (Procedure-Oriented Crew Model) is a model that includes elements of the Optimal Control Model (Kleinman et al., 1971) plus procedure execution by multiple aircrew members. It is straightforward to represent aiding for state estimation or procedure execution. It is not clear how to represent training. The simulation issues noted for KARL are relevant for PROCRU.

Fuzzy Rule-Based Model (Hunt & Rouse, 1984)

This rule-based model of troubleshooting behavior provides good predictions of sequences of behavior if an appropriate knowledge base is encoded. To an extent, knowledge acquisition is easier than with KARL since an internalized model of the system dynamics usually does not dominate maintenance performance in the same way as it affects operations. Aiding and training are potentially natural addenda - thus far, the impact of a variety of aiding concepts has been evaluated. The simulation issues noted for KARL are relevant for this model.

PROFILE model (Towne et al., 1982)

This optimization-oriented model of troubleshooting behavior can provide good predictions of expert performance. Predictions of behavioral sequences are not as good, especially for less-than-expert troubleshooters. Aiding can be represented to the extent that it affects the evolution of the feasible set. It may be possible to represent training if its effects can be captured in terms of perceptions of the feasible set. The simulation issues noted for KARL are relevant for PROFILE.

Baseline systems

The past/present predicts the future best when the future is not too different from the past/present. Consequently, the baseline model approach to predicting performance presents difficulties when new concepts and/or new technologies are of primary concern. Resolution using the baseline approach is open to choice; accuracy depends on whether or not the baseline is appropriate.

HOS Model (Lane et al., 1981)

HOS (Human Operator Simulator) is a model or modeling package that requires considerable detailed knowledge of the tasks operators are to perform. If this level of detail is available and people can be assumed to follow the prescribed paths, then reasonably good predictions are possible for performance times and perhaps errors. Aiding can be represented as task changes. Event-dependent or situation-dependent triggering of aiding may be difficult to represent.

Manual Control Models (McRuer et al., 1965; Kleinman et al., 1971).

This class of models includes both traditional manual control models and more recent performance-oriented supervisory control models. Use of these models requires knowledge of what is displayed, how it is displayed, the nature of control inputs, and the relevant system dynamics. The resolution and accuracy of the model's predictions are related to this knowledge. Aiding can be represented as modifications of control/display loops (e.g., via predictor displays). Training is much more difficult, particularly since the concern is more related to acquisition of skills, rather than knowledge. It can be difficult to represent the non-control portions of tasks (e.g., problem solving) within a control framework. Nevertheless, if these models capture the phenomena of interest, they are very powerful.

Learning Curve Models (Towill, 1989)

These models are useful for predicting performance as a function of time or number of trials. These models do *not* deal with what is learned, the process of learning, or subsequent behaviors. Accuracy depends on a comparable learning situation as a basis for extrapolation, e.g., for estimating curve parameters.

Maintainability Models (Goldman & Slattery, 1964)

This approach relies totally on task analyses and task time data, which requires much system definition. Predictions of mean time to repair depend on assumptions of particular types of distributions of task times. Training/aiding is only representable as changes of task sequences or task times.

Complexity Models (Rouse & Rouse, 1979; Wohl, 1982)

These models provide good predictions of mean time to repair as a function of system structure (i.e., number of components and the nature of interconnections). Training/aiding are representable as changes of average structural characteristics. Models of the process of dealing with complexity may eventually offer richer means of representing training/aiding.

Petri Net Model (Madni et al., 1984)

This model provides an alternative and somewhat unusual representation of troubleshooting. It is intended as a means to predict overall performance, but limited validation data is available to assess its success. It is not clear how training/aiding can be represented.

MAPPS Model (Siegel et al., 1984)

MAPPS (Maintenance Personnel Performance Simulation) is a model that places heavy emphasis on task analysis. Probabilistic branching and distributions of task times are key elements of the model. Training/aiding might be represented via modified task sequences and/or probabilities. This appears to limit the model to procedural training/aiding - which, of course, is an important alternative.

Comparability Analysis - Hardman (Weddle, 1986)

The HARDMAN (Hardware Procurement versus Military Manpower) approach depends on choosing an appropriate baseline system. High resolution and high accuracy are possible if the new system is close to the baseline, which can be a self-fulfilling prophecy. New concepts and/or new technologies can present difficulties in identifying and/or choosing appropriate baselines. Tends to be very labor intensive, but other methods, once fully developed, may require similar levels of effort.

Resource Allocation Model (Rouse, 1985)

This model provides a highly aggregated level of representation across selection, training, and several design issues. It can be used to identity high-level tradeoffs, via sensitivity analysis, but is of much less help for resolving these tradeoffs. It is difficult and expensive to obtain real data for model parameters and to tailor the model to specific applications - however, it is not inherently more difficult/expensive than comparability analysis.

Integrated Support System Tradeoff Model (Rouse, 1987)

This model is more targeted than the resource allocation model. Its current representation of training/aiding is very elementary. Most necessary data appear to be readily collectible; although, learning and retention may prove difficult to represent other than probabilistically. If substantially refined, this model might eventually be calibrated with real data and then used to produce more context-tailored guidelines.

Summary

This section has provided a very terse summary and evaluation of 29 approaches to analyzing tradeoffs between training and aiding. It should be clear by now that there is no single best approach. The choice depends on the nature of the training/aiding problem being addressed. Further, it is quite likely that success will depend on a composite approach that draws on the strengths, and compensates for the weaknesses, of several approaches.

V. COMPOSITE APPROACHES

As is illustrated in later discussion of the example, realistically complex training/aiding tradeoffs cannot be addressed with a single method, tool, or model. Further, considering the decision making context discussed earlier and illustrated in Figures 1 and 2, it is clear that many factors are usually involved in resolving training/aiding tradeoffs. Consequently, it is unreasonable to designate any of the approaches discussed earlier as the best way to pursue tradeoffs. Instead, it is more appropriate to consider how approaches can be integrated.

Three integrated or composite approaches are illustrated in Figures 7, 8, and 9. The ovals in these figures represent methods, tools, or models. The rectangles represent input information and results of using the methods, etc.

Status Quo Analysis

The composite approach depicted in Figure 7 provides a rough approximation of the ways in which training/aiding tradeoffs are currently pursued. Manpower, personnel, and training (MPT) data bases denote both data bases and a variety of spreadsheet-like tools associated with these data bases (Bogner, 1988). Man-machine systems (MMS)

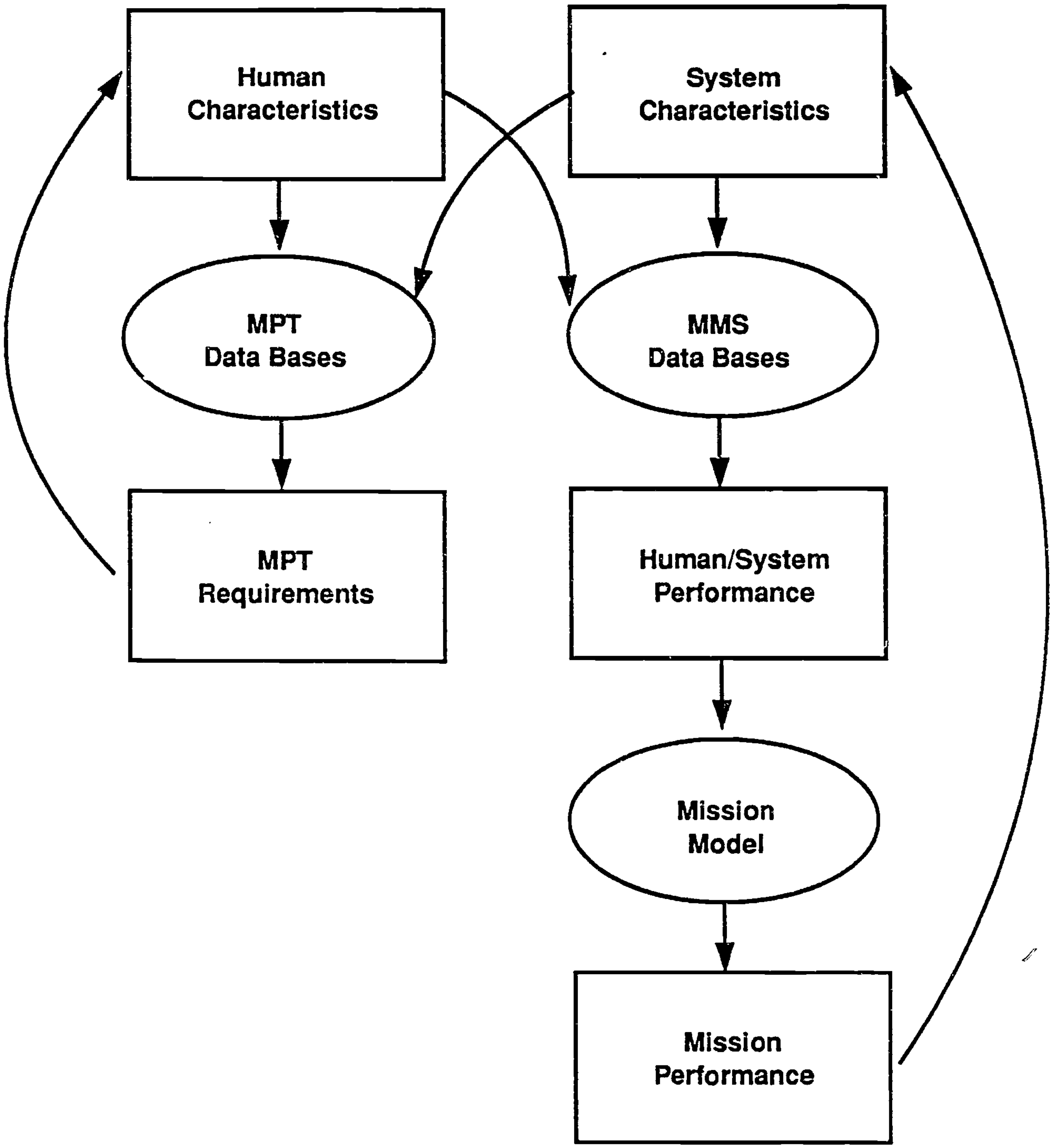

Figure 7. Status Quo Analysis.

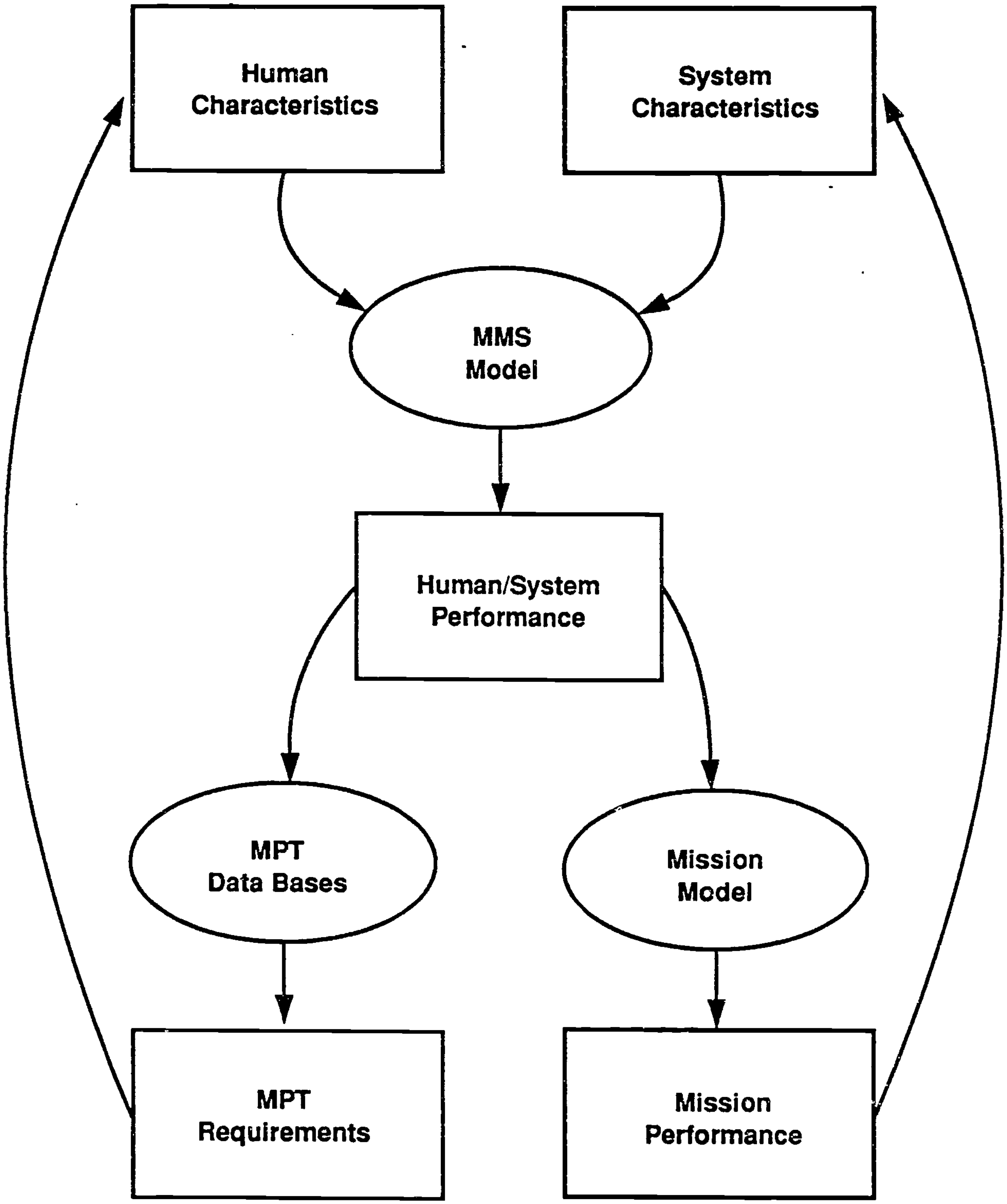

Figure 8. Performance-Based Analysis.

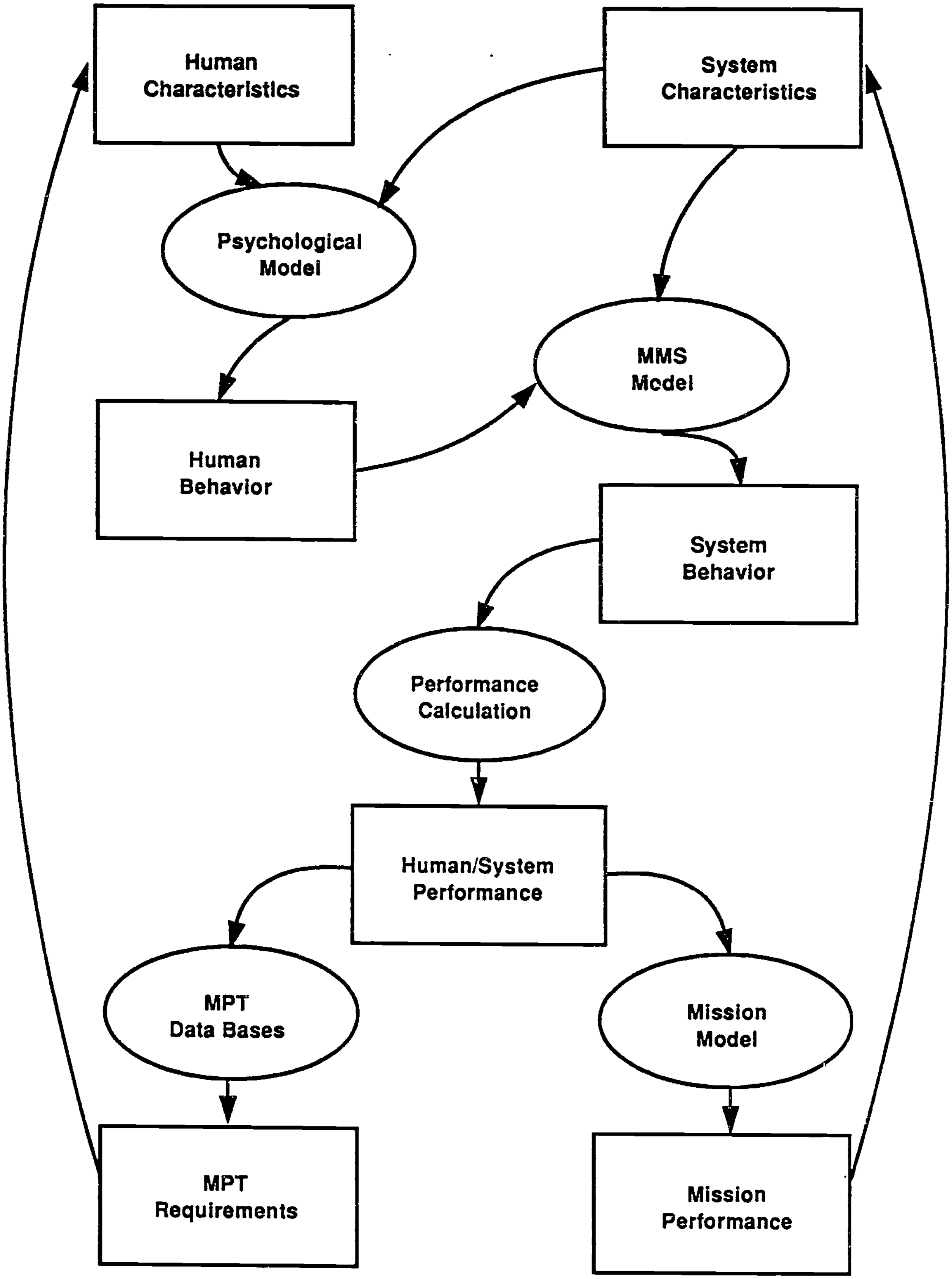

Figure 9. Behavior/Performance-Based Analysis.

data bases denote both formal and informal compilations of past experiences and experiments. Mission models typically involve computer simulations of the interactions of many people and equipment systems, as well as some representation of the mission environment. The types of model discussed in this paper potentially could provide inputs to mission models - for example, human performance parameters.

A particularly interesting aspect of Figure 7 is the relative independence of the feedback loops. If mission performance is unacceptable, system characteristics will tend to be modified independent of the impact on MPT requirements. Similarly, if MPT requirements are excessive, human characteristics will tend to be modified (perhaps via selection and classification criteria) without direct knowledge of the impact on mission performance.

As a result of the structure of Figure 7, MPT analysts and MMS designers have relatively little in common. They do not share any methods, tools, and models. Further, they often come from different disciplines (i.e., psychology and engineering) and consequently employ different concepts, jargon, etc. From this perspective, it is not surprising that systems emerge with latent MPT problems that lead to poor mission performance and/or costly redesign.

Performance-Based Analysis

Figure 8 employs a computational man-machine system model to integrate the MPT/MMS design process. In the training/aiding context, this model or, more likely, set of models is concerned with human/system performance with various levels of aiding and alternative levels of training. Later discussion of the example provides a detailed illustration of this type of use of models.

With a performance-based analysis, MPT analysts and MMS designers have to communicate in order to proceed. In a sense, the MMS model (or models) provides a unifying metaphor which both types of individual utilize and influence in somewhat different ways. Beyond facilitating communication, this approach can be the basis for providing powerful computational support for pursuing tradeoffs.

Behavior/Performance-Based Analysis

An important limitation of the performance-based approach to analysis is the possible need to go beyond performance predictions and study the behaviors underlying performance measures. This can be particularly important when the impact of training is of concern.

Training can be defined as a process of managing people's experiences so that they gain the knowledge and skills which give them the potential to perform. If one is concerned with the extent to which a particular approach to training results in acquisition of the requisite knowledge and skills, then it may be necessary to examine the process of learning as it is affected by the training regime of interest. This potentially can be accomplished by using a computational model that, via simulation, experiences the training and acquires knowledge and skills. While the state of the art is such that this approach is not feasible for most complex tasks, it is feasible for some important task components - this is illustrated in the example.

The potential use of behavioral simulations is illustrated in Figure 9. With this approach, rather than predicting performance, behavior is predicted and performance is calculated. As noted during the discussion of the evaluations of the models in the left column of Table 3, use of behavioral simulations leads to a variety of issues such as representativeness of scenarios, number of runs, appropriate statistics, etc. Of course, these issues are not new - they are similar to the issues involved in using human-in-the-loop simulators for experimental studies.

The relative lack of availability of credible and useful psychological and MMS models limits the range of applicability of the behavior/performance-based approach. Nevertheless, a few reasonable possibilities exist - see the example. Further, the practical need for these types of model may motivate support and pursuit of the research necessary to develop them.

Summary

Behavior/performance-based analyses are potentially the most powerful of the three types of analysis. Further, the outputs of behavior/performance-based analyses can be used, in principle, to produce the types of results yielded by the other two types of analysis. However, the state of the art is such that only a few viable behavior-based models are available.

In the majority of cases where such models are unavailable, performance-based analyses are a good alternative. Many models are available to support such analyses, although they must be drawn from a wide variety of disciplines. This paper represents an effort to integrate this range of material in a convincing manner. This will hopefully lead to status quo analyses being the exception rather than the rule.

VI. EXAMPLE ANALYSIS

We have thus far reviewed and integrated much material drawn from a wide range of sources. Much of the discussion has been quite conceptual and abstract. We now return to the example introduced earlier. This example provides a concrete illustration of how the methods, tools, and models reviewed earlier can be applied to a realistically complex problem.

Reviewing the earlier discussion of the example, the problem of interest is designing a head-up display (HUD) to enable long-haul truck drivers to stay on the road, avoid obstacles, and maintain speed, even though rain, fog, or snow have substantially reduced visibility. Further, this display aiding concept is to be implemented in a way that results in minimal downtime for maintenance and, in particular, avoids rerouting for specialized maintenance.

Mission Requirements

Discussions with the client and further analysis led to agreement regarding the following mission requirements:

1. Increase speed in all weather conditions - a minimum of 40 mph and a maximum of 80 mph should be attainable.
2. Maintain/improve safety - alerting times for obstacles should be at least 10 seconds and moving obstacles should be differentiated from stationary obstacles.
3. Redundant alerting messages - in the event of a failure of the HUD display unit, alerting messages should be displayed on the dashboard and/or auditorily.
4. Engine and vehicle information should also be displayed on the HUD.
5. Loss of availability for maintenance should be minimized.
6. Cost per truck should not exceed $10,000.

Initial Configuration

Consideration of the above requirements and available technology led to an initial configuration involving four units for sensors, sensor control, electronics, and display. The primary features of this configuration included:

1. Obstacle and vehicle data displayed head-up via an LCD (liquid crystal display) reflective image source.
2. Non-imaging sensor suite including:
   - Fixed beam range-only doppler radar for distant objects (800 yards)
   - Two infrared arrays (3x50) for forward sensing of angular position with 1 degree azimuth resolution over a 40 degree field of view.
   - Infrared array triangulation for near objects (<100 yards).
   - Limited elevation resolution
3. Sensor mode control based on weather, including clear, rain, fog, and snow, as well as day and night.

4. Voice synthesized audio messages in addition to HUD and dashboard display.
5. Obstacle track information maintained for positions and rates, as well as track predictions.

This initial configuration is technically feasible within the requirements and constraints listed earlier. The next question is whether or not this design is operable by truck drivers, including training and aiding implications.

Operations/Maintenance Tasks

Interacting with the HUD involves four primary operations tasks:

1. Situation interpretation - deciding which, if any, of the objects displayed on the HUD represent threats to safety.
2. Maneuver selection - choosing among alternative avoidance maneuvers, including the possibility of not maneuvering.
3. Execution and monitoring - executing the chosen maneuver and monitoring its success.
4. HUD operation - selecting modes, performing system tests, etc.

In order to satisfy the minimal downtime requirement, it is necessary for the truck driver to perform the following maintenance tasks:

1. Test verification - checking and interpreting results of system tests.
2. Fault isolation - isolating failure to lowest replaceable unit.
3. Repair decision - deciding whether or not to attempt repair.
4. Replacing units and boards - removing and replacing the lowest replaceable unit.
5. Degraded mode assessment - determining remaining functionality and the operational implications.

Initial Analysis

These nine tasks were analyzed in terms of the likely limitations, abilities, and preferences of the personnel typical for long-haul truck operations. The alternative implications of these assessments for training and aiding were then determined.

Situation Interpretation

The raw returns displayed on the HUD are likely to be overwhelming, particularly in heavy traffic situations. Some level of filtering (e.g., velocity filtering) would help. The operator would need to know how and when to adjust the filter parameters. This task could probably be proceduralized, although the assessment of when to adjust parameters has some subtleties, e.g., discriminating wet fog from light rain.

The classification of returns into threat/no threat may also be difficult. Aided classification is possible, but there inevitably will be false alarms. If aiding is needed, it will be necessary to determine the level of understanding of the classifier required for the operator to deal with the false alarms.

Maneuver Selection

Drivers might choose inefficient or inappropriate avoidance maneuvers. Computation of optimal maneuvers is feasible, but it depends on correct classification of objects and interpretation of threat's intentions. If aiding is required, it will be necessary to assure that maneuvers are executable by drivers. Also of concern will be the extent to which drivers will have to understand the optimal computation to assess the appropriateness of the maneuver.

Execution and Monitoring

Drivers' manual control abilities may be inadequate for some of the necessary maneuvers. Automation is feasible, but the resulting monitoring task might be difficult (but far from boring) and driver acceptance is likely to be quite low. Further, such automation could result in exceeding the $10,000 per truck constraint. If aiding is required, it will be necessary to determine how the driver is to decide the automation has failed, and how automatic to manual transitions will occur. Some type of simulator training may be needed.

HUD Operation

Driver has to know why, how, and when for mode selection. Procedures would be of use and perhaps could be embedded in the system. HUD system might also provide feedback to driver on the implications of mode selections, for example, in terms of likely sensor performance.

Test Verification

A simple "red light" may be inadequate if the driver is to perform any maintenance. Depending on what functionality remains, the system could provide online explanations and embedded training. This material could, of course, also appear in hardcopy system manuals.

Fault Isolation

Built-in test could do most of this task, but probably not all of it. With some diagnostic aiding, the driver might be able to replace boards rather than whole units. Embedded training might be part of the diagnostic aiding.

Repair Decision

The driver might have difficulty with deciding whether or not to attempt repair. This decision is affected by the degraded mode assessment (see below) and the availability of spares. Repair decisions could probably be proceduralized, perhaps supported by embedded training.

Replacing Units and Boards

This task should be straightforward and easily proceduralized. Embedded procedures and the hardcopy manuals could provide support. Minimal training is likely to be needed.

Degraded Mode Assessment

It may be difficult for the driver to determine the remaining functionality and the operational implications. Online aiding may not be feasible. A central issue is the knowledge requirements for driver to be able to generate assessments and understand implications.

Primary Tradeoffs

The above analysis led to identification of 10 aiding alternatives and one or more training implications of each alternative. The value of using procedures and associated training was obvious for several tasks and no further analysis was pursued. Also, in two cases, resolution of training/aiding issues interacted across tasks. As a result of these considerations, the overall analysis was reduced to four primary tradeoffs.

Object Classification

The performance metrics of interest for this task are threat identification time and classification errors (both misses and false alarms). The tradeoff is between unaided human performance and aided classification where the human must detect the false alarms of the aid.

Human performance in this task could predicted using Greenstein's pattern recognition model of event detection (Greenstein & Rouse, 1982)[2]. Learning time for this task could be estimated by using data from Boff and Lincoln's *Compendium* (1988; entries 4.201 and 7.414-7.416) or modeled using Rouse's model for learning stochastic estimation tasks (Rouse, 1977). Drivers' abilities to detect aiding failures could also be modeled in the same way.

The inputs to these models would be displayed features and decision criteria. The outputs would be response time and errors as a function of number of targets and threat density. In order to choose between unaided and aided, one would have to know the relative values of performance (i.e., time versus errors) and the life-cycle costs of aiding and training alternatives. These values and costs might be determined by using the outputs of the models noted here as inputs to a broader long-haul trucking mission model.

Preview Control

This tradeoff involves both maneuver selection and execution. The performance measures are path "optimality" and root-mean-squared (RMS) control errors. The tradeoff is between unaided human performance and various levels of automation which the driver must monitor.

Human performance in this task could be predicted using Govindaraj's model of preview control (Govindaraj & Rouse, 1981), with enhancements for path selection[2]. Learning time for this task could be estimated using data from Boff and Lincoln's *Compendium* (1988; entries 4.201, 9.402, and 9.539) or modeled using Towill's (1989) learning curves. Detection of automation failures could be modeled using the models discussed for the classification tradeoff.

---

[2] This model is a member of the manual/supervisory control class of models in Table 3.

The inputs to the preview control model would be a representation of the vehicle dynamics and the path error criteria. Outputs would be RMS errors as a function of path characteristics and number of threats. To choose between unaided performance and various levels of automation, one would have to know the value of RMS error relative to broader vehicle system metrics and the life-cycle costs of aiding and training. The truck mission model discussed earlier might be a source of this information.

Troubleshooting

This tradeoff involves both test verification and fault isolation. Performance measures are diagnostic time and errors. The tradeoff is between unaided and aided diagnosis, including the training implications of both alternatives.

Human performance in this task (with and without aiding) could be modeled using Hunt's fuzzy, rule-based model of troubleshooting performance (Hunt & Rouse, 1984). Inputs to the model include the context-specific and context-free knowledge bases (rules) of the driver. Outputs would be diagnostic time and errors as a function of average initial feasible set size.

While this model can, in principal, learn new rules as the result of training, it has not actually been tried. An alternative approach, however, is to use the model to determine knowledge requirements to satisfy minimal downtime objectives. These knowledge requirements can serve as basis for determining training requirements, costs, etc. The overall tradeoff could be resolved by using the unaided and aided mean time to repair, diagnostic errors, training times, etc. as inputs to a trucking logistics model.

Degraded Mode Assessment

The performance measures of interest here are decision times and decision errors. The tradeoff is between various levels of training for unaided human performance.

Performance in this task could be predicted using Knaeuper's rule-based model of planning and decision making in the control of dynamic processes (Knaeuper & Rouse, 1985). Inputs to this model would include a representation of the vehicle dynamics and the HUD functionality, as well as the knowledge base of formal and informal operating procedures for the truck and HUD. Outputs would be the frequencies of incorrect assessment decisions.

As noted in the earlier discussion, this model can, in principle, learn new rules, but this has not actually been tried. Consequently, it would be more appropriate to use this model to determine knowledge requirements to achieve acceptably low error frequencies. These requirements could then drive a training analysis. The tradeoff could ultimately be resolved using some combination of the mission and logistics models.

Summary

This example has served to illustrate how the methodological framework presented in this paper can be applied to a realistically complex problem. Clearly, what started as a fairly straightforward problem, soon became relatively complicated. Nevertheless, there are methods available to address these complications. In fact, the models we chose were just a few of many possibilities - we chose the ones with which we were most familiar.

This example also portrays how judgment-laden such an analysis can be. It is rather difficult to imagine proceduralizing the analysis, other than in a skeletal form to organize one's thinking. Also, the tradeoffs identified in this analysis could not be resolved without broader knowledge of the truck's mission, support system, and likely scenarios.

The role of judgment, as well as the complications and complexity, lead to a question of how the framework presented in this paper can be supported. How can an analyst manage so many issues, methods, and analyses? How can the wealth of information needed be readily accessed?

These questions should be addressed in an incremental manner. First, the framework should be refined by applying it to several additional examples. These experiences will help to define a fuller set of methods, tools, models, and data sources to be incorporated in the framework. These results will, in turn, provide a basis for scoping a computer-based system for supporting use of the framework. The potential nature of such a computer-based system is discussed in the next section.

VII. CONCLUSIONS

This paper covers a lot of ground. This breadth is required to address the tradeoff between putting "smarts" in people versus putting "smarts" in machines. This issue cannot be resolved appropriately in isolation from the broader issues of mission requirements and MPT implications.

Despite this complexity, this paper has shown that the issue can be addressed in other than an ad hoc manner. A wide range of methods, tools, and models are applicable to analysis of various aspects of training/aiding tradeoffs. Thus, these tradeoffs can be rigorously formulated in principle - but, does the state of the art let us pursue such formulations in practice?

The answer to this question hinges on two types of issue. Research issues refer to the need to integrate our knowledge of training and aiding to enable early model-based predictions of the impact of design decisions. Integration issues include the set of problems associated with packaging methods, tools, models, data sources, and support "utilities" into a coherent approach to supporting the types of analysis described in this paper.

Research Issues

Progress is needed in several areas before the framework presented in this paper will have widespread practical utility. In general, we need to move beyond measurement and classification of behavioral phenomena to develop predictive models that will enable tradeoff studies prior to producing a detailed design for an equipment system. This will enable consideration of training/aiding issues, as well as broader MPT issues, during the early states of conceptual design rather than the later stages of detailed design. An important implication of such early involvement will be the need for models that do not rely on information from traditional task analyses - the HUD example shows how one can proceed with the formulation of tradeoffs without detailed task analyses.

The ability to perform training versus aiding tradeoffs prior to detailed design is of great value. With this ability, MPT issues and human factors issues associated with aiding can be considered quite early, and design decisions can be changed with less impact on schedule and cost. Thus, for example, our analyses have shown that truck drivers may have difficulty with degraded mode assessments. This potential difficulty has important implications for either increased driver training or greater intelligence built into the system. Model-based analyses of such tradeoffs can guide the development of conceptual designs, rather than waiting to react to detailed designs.

The need for predictive models is likely to be satisfied more easily when the goal is predicting the impact of aiding on human performance. This paper has discussed a reasonably impressive range of models suitable for this purpose. In contrast, predictive models of the impact of training are rare. While learning curve models are of some value, they do not address the question of what is learned as a function of varying levels of training. To satisfy this objective, effort needs to be invested in developing computational models of the learning processes underlying training of operations and maintenance tasks.

We also need to apply existing and emerging models to analysis of existing training and aiding alternatives. The literature is replete with reports of studies of training and aiding. Thus, for example, we know that feedback and explanations about errors usually enhances troubleshooting training and that predictor displays often improve manual control performance. For tradeoff analyses, however, we also need to be able to predict (as opposed to measure) the impact of alternative training and aiding notions in terms of quantitative performance metrics - at the very least, we need to be able to make predictions of relative differences. To provide this capability, we need to integrate understanding of existing training and aiding alternatives within a model-based framework such as presented in this paper.

Looking beyond "traditional" training and aiding concepts, few, if any, of the available models deal explicitly with the impact on human performance of intelligent training and/or aiding systems. Intelligent systems react to human performance and adapt the nature of the interaction accordingly. A variety of such systems are under development which integrate training and aiding within a single support system (Johnson, 1981; Keskey & Sikes, 1987; Kline & Lester, 1988; Richardson et al., 1986). The design philosophies and software architectures underlying these systems reflect training/aiding tradeoffs, albeit usually implicitly. The framework presented in this paper could provide the basis for making these tradeoffs systematically and rigorously. A critical issue for such applications is the ability to model humans' performance in a system that adapts to their performance - of course, it is also quite likely that people will adapt to the adaptive system!

Integration Issues

Beyond developing better predictive models, and focusing special attention on learning processes and intelligent systems, a very important issue concerns how the range of methods, tools, and models discussed in this paper can be utilized in a consistent and comprehensive manner. How can one or more analysts access and utilize such a wide range of approaches, including supporting data and documentation?

The obvious answer is to develop a computer-based version of the framework presented in this paper. However, as useful as such a computer package might be, we expect that more "value added" will be needed to have this framework widely adopted. This might be accomplished by developing an "analysis environment" that not only accesses the information in this paper, but also provides facilities for creating, manipulating, and linking alternative representations. Also of value would be a variety of "utilities" for supporting many analysis activities beyond accessing and utilizing methods, tools, and models.

The detailed requirements for such an "environment" could be determined by using the framework presented here to pursue several additional examples. With a sufficiently broad range of examples, requirements for methods, tools, models, and data sources can be refined. Further, study of the course of these additional anlyses will enable identification of utilities that should be included within the analysis environment. These results could provide the basis for a prototype environment, use of which could be evaluated with a range of users. Such an evaluation could assess users' reactions to the concept, how their analysis behaviors are affected, and the performance impact of working in the environment.

An analysis environment could be very useful long before it was complete. The overall framework, currently available models, and several straightforward support utilities could substantially improve the ways in which training/aiding tradeoffs are currently resolved. As the results of the aforementioned research efforts emerge, they could be embedded in the environment. In the process, an evolving and increasingly powerful toolbox would enable us to make smart decisions about how much smarts to put in people and how much to put in machines.

REFERENCES

Akman Associates. (1987). *RAMPARTS handbook: A blueprint for enhanced MPTS planning.* Silver Spring, MD: Akman Associates.

Allen, T.J. (1977). *Managing the flow of technology.* Cambridge, MA: MIT Press.

Baron, S. (1984). A control theoretic approach to modeling human supervisory control of dynamic systems. In W.B. Rouse, (Ed.), *Advances in man-machine systems research* (Vol. 1, pp. 1-47). Greenwich, CT: JAI Press.

Baron, S., & Kruser, D.S. (Eds.). (1988). *Human performance modeling.* Washington, DC: National Academy Press.

Boff, K.R., Kaufman, L., & Thomas, J.P. (1986). *Handbook of perception and human performance.* New York: Wiley.

Boff, K.R., & Lincoln, J.E. (1988). *Engineering data compendium: Human perception and performance.* Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.

Bogner, M.S. (1988). *Catalogue of MANPRINT methods.* Alexandria, VA: U.S. Army Research Institute.

Booher, H.R. (1978). *Job performance aid selection algorithm: Development and application* (Rept. TN-79-1). San Diego, CA: Navy Personnel Research and Development Center.

Booz-Allen, & Hamilton. (1985). *Air Force manpower, personnel, and training (MPT) systems model course: MPT systems model student textbook.* Bethesda, MD: Booz-Allen & Hamilton, Inc.

Cody, W.J., & Rouse, W.B. (1989). A test of criteria used to select human performance models. In G. R. McMillan (Ed.), *Applications of human performance models to system design.* New York: Plenum Press.

Elkind, J.W., Card, S.K., Hochberg, J., & Huey, B.M. (Eds.). (1989). *Human performance models for computer-aided engineering.* Washington, DC: National Academy Press.

Fleger, S.A., Permenter, K.E., & Malone, T.B. (1987). *Advanced human factors engineering tool technologies.* Fairfax, VA: Carlow Associates.

Foley, J.P., Jr. (1978). *Executive summary concerning the impact of advanced maintenance data and task oriented training technologies on maintenance, personnel, and training systems* (AFHRL-TR-78-24 AD-A053 682). Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory.

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. In M.R. Rosenzweig & L.W. Porter, (Eds.), *Annual review of psychology* (Vol. 40, pp. 631-666). Palo Alto, CA: Annual Reviews, Inc.

Goldman, A.S., & Slattery, T.B. (1964). *Maintainability.* New York: Wiley.

Govindaraj, T., & Rouse, W.B. (1981). Modeling the human controller in environments that include continuous and discrete tasks. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-11*, 410-417.

Greeno, J.G., & Simon, H.A. (1988). Problem solving and reasoning. In R.C. Atkinson, R. Hernstein, G. Lindzey, & R.D. Luce, (Eds.), *Stevens' handbook of experimental psychology* (rev. ed.). New York: Wiley.

Greenstein, J.S., & Rouse, W.B. (1982). A model of human decision making in multiple process monitoring situations. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-12*, 182-193.

Hunt, R.M., & Rouse, W.B. (1984). A fuzzy rule-based model of human problem solving. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-14*, 112-120.

Irvin, J.G., Blunt, J.H., & Lamb, T.A. (1988). *A training/JPA model: Its utility for the MPT community*. Dayton, OH: Universal Energy Systems, Inc.

Johnson, R.C. (1981). *Integrated maintenance information system: An imaginary preview* (AFHRL-TP-81-18 AD-A104 224). Wright-Patterson AFB, OH: Logistics and Technical Training Division, Air Force Human Resources Laboratory.

Keskey, L.C., & Sikes, D.J. (1987). An artificial intelligence (AI)- simulation based approach for aircraft maintenance training. *Proceedings of the Society for Computer Simulation Simulators Conference*, 285-290.

Kleinman, D.L., Baron, S., & Levison, W.H. (1971). A control theoretic approach to manned-vehicle system analysis. *IEEE Transactions on Automatic Control, AC-16*, 824-832.

Kline, C.R., Jr., & Lester, M. (1988). *A model for integrating a job-aiding, training, and performance assessment system--A preliminary concept paper* (AFHRL-TP-87-60 AD-A192 473). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Knaeuper, A.E., & Rouse, W.B. (1985). A model of human problem solving in dynamic environments. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-15*, 708-719.

Kulik, J.A., & Kulik, C-L.C. (1988). *Meta-Analysis in educational research*. Ann Arbor, MI: University of Michigan, Center for Research on Learning and Teaching.

Lane, N.E., Strieb, M.I., Glenn, F.A., & Wherry, F.J. (1981). The human operator simulator: An overview. In J. Moraal & K.F. Kraiss (Eds.), *Manned system design: Methods, equipment, and applications* (pp. 121-152). New York: Plenum Press.

McMillan, G.R. (Ed.). (1989). *Applications of human performance models to system design*. New York: Plenum Press.

McRuer, D.T., Graham, D., Krendel, E.S., & Reisener, S. Jr. (1965). *Human pilot dynamics in compensatory systems - Theory, models, and experiments with controlled element and forcing function variations* (Tech. Rept. AFFDL-TR-65-15). Wright-Patterson AFB, OH: Air Force Flight Dynamics Laboratory.

Madni, A.M., Chu, Y.Y., Purcell, D., & Brenner, M.A. (1984). *Design for maintainability with modified petri nets* (Tech. Rept. PFTR-1125-84-6). Woodland Hills, CA: Perceptronics, Inc.

Moraal, J., & Kraiss, K.F. (Eds.) (1981). *Manned systems design: Methods, equipment, and applications.* New York: Plenum Press.

Morris, N.M., & Rouse, W.B. (1985). Review and evaluation of empirical research in troubleshooting. *Human Factors, 27,* 503-530.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering.* New York: North Holland.

Richardson, J.J., Anselme, C.J., Harmon, K.R., Keller, R.A., & Moul, B.L. (1986). *Artificial intelligence technology for the Maintainer's Associate* (AFHRL-TR-86-31 AD-A175 211). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Rossmeissl, P.G., Alderson, M.A., Shotzberger, L.J., & Payne, D.L. (1989). *The cognitive requirements model: An approach to job engineering.* Washington, DC: HAY Systems, Inc.

Rouse, W.B. (1977). A theory of human decision making in stochastic estimation tasks. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-7,* 274-283.

Rouse, W.B. (1980). *Systems engineering models of human-machine interaction.* New York: North Holland.

Rouse, W.B. (1985). Optional allocation of system development resources to reduce and/or tolerate human error. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-15,* 620-630.

Rouse, W.B. (1987). Model-based evaluation of an integrated support system concept. *Large-Scale Systems, 13,* 33-42.

Rouse, W.B., & Cody, W.J. (1988). On the design of man-machine systems: Principles, practices, and prospects. *Automatica, 24,* 227-238.

Rouse, W.B., & Cody, W.J. (1989a). Designers' criteria for choosing human performance models. In G.R. McMillan (Ed.), *Applications of human performance models to system design.* New York: Plenum Press.

Rouse, W.B., & Cody, W.J. (1989b). Information systems for design support: An approach for establishing functional requirements. *Information and Decision Technologies,* in press.

Rouse, W.B., Hammer, J.M., & Lewis, C.M. (1989). On capturing humans' skills and knowledge: Algorithmic approaches to model identification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-19.*

65

Rouse, W.B., & Rouse, S.H. (1979). Measures of complexity in fault diagnosis tasks. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-9*, 720-727.

Sheridan, T.B., & Ferrell, W.R. (1974). *Man-machine systems: Information, control, and decision models of human performance.* Cambridge, MA: MIT Press.

Siegel, A.I., Bartter, W.D., Wolf, J.J., & Knee, H.E. (1984). *Maintenance personnel performance simulation (MAPPS)* model (Tech. Rept. ORNL/TM-9041). Oak Ridge, TN: Oak Ridge National Laboratory.

Smillie, R.J., & Blanchard, R.E. (1986). Job performance aids in an integrated personnel system. In J. Zeidner (Ed.), *Human productivity enhancement* (Vol.1, pp. 167-199). New York: Praeger.

Thurman, M.R. (1989). Analysis counts. *Bulletin of Military Operations Research, 22,* 4-8.

Towill, D.R. (1989). Selecting learning curve models for human operator performance. In G.R. McMillan (Ed.), *Applications of human performance models to system design.* New York: Plenum Press.

Towne, D.M., Johnson, M.C., & Corwin, W.H. (1982). *PROFILE: A technique for projecting maintenance performance from design characteristics* (Tech. Rept. 100). Los Angeles, CA: University of Southern California, Behavioral Technology Laboratories.

Wasserman, A.I., & Shewmake, D.T. (1985). The role of prototypes in the user software engineering (USE) methodology. In H.R. Hartson (Ed.), *Advances in human-computer interaction* (pp. 191-209). Norwood, NJ: Ablex.

Weddle, P.D. (1986). Applied methods for human-systems integration. In J. Zeidner (Ed.), *Human productivity enhancement.* (Vol. 1, pp. 332-363). New York: Praeger.

Wohl, J.G. (1982). Maintainability prediction revisited: Diagnostic behavior, system complexity, and repair time. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-12*, 241-250.

66