Technical Report 674

Comparison-Based Predictions and Recommendations for Army Maintenance Training Devices

AD-A170 942

Gary A. Klein and Sallie E. Gordon
Klein Associates

Mark Palmisano and Angelo Mirabella
Army Research Institute

Training and Simulation Technical Area

Training Research Laboratory

ari

U. S. Army

Research Institute for the Behavioral and Social Sciences

March 1985

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

L. NEALE COSBY
Colonel, IN
Commander

Research accomplished under contract
for the Department of the Army

Klein Associates

Technical review by

V. Irizarry
J. Kaplan

NOTICES

~~DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POT, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600~~

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

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Technical Report 674	2. GOVT ACCESSION NO. AD-A170942	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COMPARISON-BASED PREDICTIONS AND RECOMMENDATIONS FOR ARMY MAINTENANCE TRAINING DEVICES		5. TYPE OF REPORT & PERIOD COVERED Interim Report April 1983-April 1985
		6. PERFORMING ORG. REPORT NUMBER TR-84-41-7
7. AUTHOR(s) Klein, G. A., & Gordon, S. E. (Klein Associates); & Palmisano, M., & Mirabella, A. (ARI)		8. CONTRACT OR GRANT NUMBER(s) MDA 903-C-83-0270
9. PERFORMING ORGANIZATION NAME AND ADDRESS Klein Associates 740 Wright Street Yeilow Springs, OH 45387		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q162717A790 3413 102
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		12. REPORT DATE March 1985
		13. NUMBER OF PAGES 69
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) --		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE --

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

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

--

18. SUPPLEMENTARY NOTES

Contracting Officer's Representative, Ray Perez.

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

Maintenance training devices
Generic maintenance training devices
Instructional features assessment
Hands-on practice
Skill maintenance

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

This study is part of a larger project to evaluate the use of Comparison-Based Prediction (CBP) techniques as a method to estimate the effectiveness of new training devices and concepts, and to generate design recommendations.

The goal of this effort was to use CBP to predict the savings in training time that might be realized by introducing maintenance training devices into automotive courses for Army personnel. The intent was to learn about the strengths and weaknesses of the method when used by Army instructional personnel.

(Continued)

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

UNCLASSIFIED

 SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ARI Technical Report 674

20. (Continued)

A secondary goal was to use CBP to generate recommendations for improving the design of maintenance training devices.

Predictions were generated separately for two prototype training devices, one developed by Grumman and one developed by Seville/Burtek for use in automotive courses at Aberdeen Proving Ground and Edgewood Arsenal. The approach used Subject Matter Experts (SMEs) to (1) identify appropriate comparison cases to be used to provide baseline data, (2) estimate differences between the prototype and the comparison cases, and (3) generate predictions by adjusting the baseline data to account for the differences between the prototypes and comparison cases. Subject Matter Experts compared the two prototype training devices either to actual equipment trainers or to maintenance simulators procured for the Air Force.

The results showed that CBP can be used for training device development. The Army instructional personnel were able to work with this method, and an important by-product was a set of opinions about effective and ineffective device features.

Technical Report 674

Comparison-Based Predictions and Recommendations for Army Maintenance Training Devices

Gary A. Klein and Sallie E. Gordon
Klein Associates

Mark Palmisano and Angelo Mirabella
Army Research Institute

Contracting Officer's Representative
Ray Perez

Training and Simulation Technical Area
Robert J. Seidel, Chief

Training Research Laboratory
Donald F. Haggard, Acting Director

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

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

March 1985

Army Project Number 2Q162717A790

Training Devices and Simulation

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

FOREWORD

An ARI mission is to produce technology (i.e., job aids) that will help Army training developers design, acquire, and use simulation-based and computer-based programs of instruction, particularly for automotive and electronic maintenance training. A critically needed type of aid is one that will help design and evaluate training devices early in the weapon acquisition cycle.

One approach to such aiding--comparison-based prediction--is the subject of this report. The approach has been used successfully as part of the HARDMAN method for estimating new hardware reliability. The current research attempts to exploit that methodology for estimating the effectiveness of training devices as early as the drawing board or prototype stage of training development. The results, though preliminary, are encouraging. They provide part of the basis for preparing user-oriented guidelines that will emerge as the end-product of this project. These guidelines should help training device procurers such as the PM-TRADE and training developers in the U.S. Army Training and Doctrine Command and also the Army Research Institute.

EDGAR M. JOHNSON
Technical Director

ACKNOWLEDGMENTS

This report depended on the support of many people. At the Army Research Institute, Bob Hays and Mike Singer made important suggestions about data collection strategies; and Jonathan Kaplan's criticisms and questions helped focus and clarify our report. Dan Tufano should also be acknowledged for helping structure some of the early planning. At Aberdeen Proving Ground, Jim Dees again made sure that we were able to conduct our data collection interviews. We also wish to thank Bob Swezey, Joe Moyer, and Martha Cole of SAI for all of the support we received in understanding the AMTESS devices and preparing our materials. In addition, Paula John provided valuable editorial assistance.

Comparison-Based Predictions and Recommendations for Army Maintenance Training Devices

EXECUTIVE SUMMARY

Requirement:

This study is part of a larger project to evaluate the use of Comparison-Based Prediction (CBP) techniques as a method to estimate the effectiveness of new training devices, training device concepts, and to generate design recommendations. CBP has been used successfully for logistics. This program is an attempt to study the value of CBP for training system design and evaluation.

The goal of this effort was to use CBP to predict the savings in training time that might be realized by introducing maintenance training devices into automotive courses for Army personnel. No evaluation of the accuracy of these predictions was planned for this effort. Rather, the intent was to learn about the strengths and weaknesses of the method when used by Army instructional personnel. Once the methods are sufficiently developed, a test of predictive accuracy will be needed.

A secondary goal was to use CBP to generate recommendations for improving the design of maintenance training devices.

Procedure:

Predictions were generated separately for two prototype training devices, one developed by Grumman and one developed by Seville/Burtek for use in automotive courses at Aberdeen Proving Ground and Edgewood Arsenal. The approach used Subject Matter Experts (SMEs) to (1) identify appropriate comparison cases to be used to provide baseline data for the predictions, (2) estimate differences between the prototype and the comparison cases, and (3) generate predictions for the prototypes by adjusting the baseline data to account for the differences between the prototypes and comparison cases. Two types of comparison cases were used in the study. Subject Matter Experts compared the two prototype training devices either to actual equipment trainers or to maintenance simulators procured for the Air Force.

Findings:

The results showed that CBP can be used for training device development. The Army instructional personnel were able to work with this method to provide the information needed to generate predictions. The application also identified several important issues for future research about this method of prediction. These include: (1) selecting Subject Matter Experts, (2) defining critical differences between comparison

cases, (3) preparing explicit task descriptions, and (4) validating the predictions made by the method. An important by-product was a set of opinions about effective and ineffective device features.

Comparison-based recommendations were derived for a next generation of Army automotive maintenance trainers. The recommendations covered hardware, software, instructional software, and also addressed factors related to their efficient utilization. In addition, a special recommendation was made about a new type of training device, based on the comparison of Army prototype to Air Force maintenance training devices. These recommendations suggest that comparison-based methods can serve both to evolve training device design, and to identify new device concepts.

Comparison-Based Predictions and Recommendations
for Army Maintenance Training Devices

CONTENTS

	Page
EXECUTIVE SUMMARY	vii
INTRODUCTION	1
Comparison-Based Prediction (CBP) Approach	3
The Use of CBP in Predicting Training Effectiveness of Prototype Maintenance Training Devices (TDs)	8
OBJECTIVES OF THE STUDY	10
METHOD	11
Training Device Characteristics	11
Courses and Tasks	11
CBP Strategy	13
Causal Factors	17
Subjects	18
Materials	20
Procedure	21
RESULTS: PREDICTIONS OF TD EFFECT ON COURSE TIME	21
Grumman Prototype	22
Seville/Burtek TD, Course 63W10	29
Seville/Burtek TD, Course 63B30	31
Instructor Time	35
RESULTS: CONCEPTUAL DESIGN RECOMMENDATIONS	35
Feature Comparison of Grumman vs. Seville/Burtek	35
Composite Design Recommendations	36
Summary of Design Recommendations	36
Software and Courseware	40
Instructional Support and Utilization	41
Availability and Reliability	42
Alternative TD Recommendations	42
SUMMARY, DISCUSSION, AND CONCLUSIONS	43
TD Predictions and Recommendations	43
Recommendations for Future Maintenance TDs	46
The Concept of Generic Maintenance Training Devices	46
Assessment of Comparison-Based Prediction	48

Page

REFERENCES 55

APPENDIX A 57

LIST OF TABLES

Table 1. CBP process 6

2. Advantages of comparison-based prediction 9

3. Comparative matrix of AMTESS design characteristics 12

4. Course tasks used in predicting training effectiveness 14

5. Summary of SME background 19

6. Predicted time impact for Grumman TD in Course 63D30, for SME 1 23

7. Predicted time impact for Grumman TD in Course 63H30, for SME 2 25

8. Predicted time savings from Grumman TD in Courses 63D30 and 63H30 as compared to Air Force 6883 TDs 27

9. Predicted savings for Seville/Burtek TD, for Course 63W10, for SME 3 30

10. Predicted time savings for Seville/Burtek TD (63W10 Course) as compared to Air Force 6883 Honeywell TD 32

11. Predicted savings for Seville/Burtek TD, Course 63B30, for SME 4 33

12. Feature comparison for automotive and missile tasks 37

13. Summary of predicted savings in training time for Grumman and Seville/Burtek TDs 45

14. Potential impact of causal factors on training time 50

15. Potential impact of causal factors on instructor time 51

Page

LIST OF FIGURES

Figure 1. Planning cycle 2

2. Logic of comparison-based prediction 5

3. Schematic diagram of the use of multiple comparisons 16

INTRODUCTION

This report is part of an Army Research Institute program to develop tools for estimating training device effectiveness. It is now difficult and costly to do so prior to delivery when simulator design can be modified and optimal configurations can be identified. The reason for this difficulty is that relevant data are usually not available until the later stages of development when design modification may be impractical (Figure 1).

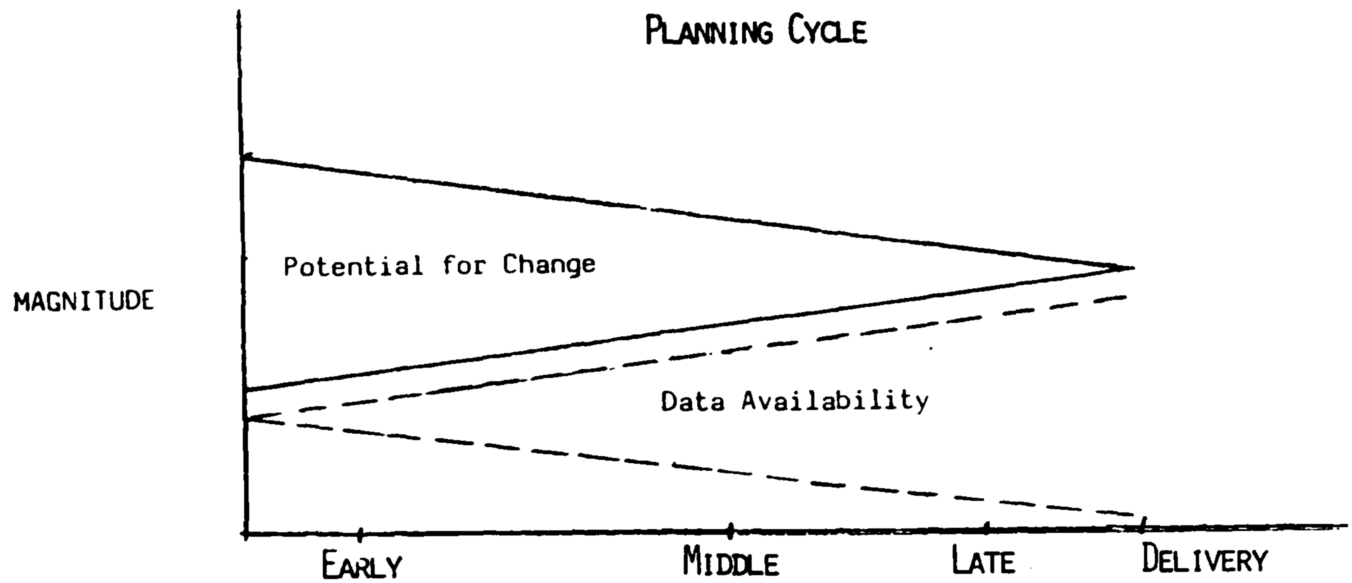
When early predictions of training device effectiveness are needed, one approach is to rely on expert judgements. These are relatively easy to obtain but somewhat more difficult to quantify. Consequently, it is hard to put much confidence in them since their rationale is usually not clear.

Another approach is to develop formal models. However, these require a great deal of data. It does not appear likely that there will be an operational data base for training device predictions in the near future, making it difficult to develop formal models for predicting training device effectiveness. Any applicable approach to prediction will probably rely on subject matter expert judgement. It is therefore important to find ways to structure and improve the quality of these judgements by making them more analytical.

Currently the only prediction technique available for this type of situation with missing or ambiguous data is TRAINVICE (Tufano 1982). TRAINVICE uses ratings of the extent to which a training device matches a task on physical and functional similarity of controls and displays, similarity of task performance, and related analyses of learning deficits and training techniques, in accordance with a variety of training principles. But TRAINVICE is limited because its predictions are based on a logically derived index of transfer of training whose scale properties are unknown. In addition it requires judgements that are labor intensive and cumbersome to apply.

An alternative that is easy to use, applicable to a wide range of situations, and face valid is Comparison-Based Prediction (CBP). CBP relies on structured interviews or questionnaires, using reasoning by analogy to compare the prediction target to existing operational cases. This approach, successfully applied to logistics predictions (McManus 1979), is being extended to predicting training device effectiveness. Experts can be asked to estimate the effectiveness of a new or proposed training device by comparing it with an existing one. Three applications have been planned: to study how CBP would have to be modified to make predictions of the effectiveness both of maintenance and of tank gunnery training devices; and to make recommendations about training programs for a new artillery system.

Figure 1



PLANNING CYCLE
MAGNITUDE
Potential for Change
Data Availability
EARLY
MIDDLE
LATE
DELIVERY

In the next section, we describe the CBP approach and then detail its application to the prediction of maintenance training device effectiveness.

Comparison-Based Prediction (CBP) Approach

CBP is reasoning by analogy, where an inference is made for one object or event based upon another similar object or event. The Air Force developed a procedure for such reasoning (Comparability Analysis) to predict the reliability of new aircraft subsystems (Tetmeyer, 1976). For example, an engineer wishing to predict the reliability of the duct system of a new aircraft could find a duct system on a comparable aircraft that was already in use. The operational data on the reliability of the existing duct system could serve as a data base. The engineer would identify differences between the new aircraft and the operational one that would affect reliability. These differences would be used to estimate an adjustment factor. If the duct system of the new aircraft was twice the size of the duct system of the existing aircraft, and there were no other important differences, then the adjustment factor might 2. The engineer could apply this adjustment factor to the operational data to generate a prediction for the new duct system. Thus the prediction would be based on the operational data but enhanced by the engineer's judgement of how to adjust those data to fit the new situation.

Klein Associates assessed the process of Comparability Analysis (Klein & Witzenfeld, 1982) and presented an explanation of the logic underlying the use of comparison cases to derive predictions (Weitzenfeld, 1984; Weitzenfeld & Klein, 1982). These investigators were interested in improving the method and increasing its range of application beyond reliability assessment and logistics. They therefore examined the logic of using that analogical inference to generate predictions.

Numerous models of analogical reasoning have been proposed by researchers, including the proportion model (a:b:c:d) examined by Sternberg (1977), the similarity matching model of Tversky (1973), and the philosophy of science model presented by Hesse (1966). However, when studying these models (Klein & Weitzenfeld 1982), found that none seemed to reflect the most important aspect of Comparability Analysis choosing an appropriate analogous situation, in order to assess the difference between the current situation, and the predicted analogue, thereby deriving an inference by making appropriate adjustments in data obtained from the comparison case. The proportion model presents subjects with the analogy, but does not reflect how people identify good comparison cases, nor does it address how they make adjustments if the analogy does not fit perfectly. The philosophy of science model focuses on the identification of new theories and laws, but not with the application of existing knowledge. The similarity matching model does not include a means of identifying relevant and important similarities, but treats all similar features as equivalent.

We have, therefore, suggested a model of analogical reasoning that emphasizes the role of causal factors (Klein, 1982; Weitzenfeld, 1984; Klein, Gordon, Palmisano, and Mirabells, 1984). This model states that for situation (A) there is a set of causal factors (x,y,z...) that will influence or determine (t), the target characteristic of (A) to be estimated.

Situation A could be a new aircraft duct system; causal factors x,y,z could be the size of the duct system, number of flying hours, type of mission; and A(t) could be the reliability of the new duct system.

In determining the target variable, A(t), we usually cannot identify all of the causal factors involved, their effects or interactions. Instead, an analogous situation or comparison case (situation B, another aircraft) is identified which reflects the same determinants as the target cases (see Figure 2). That is, for situation B, the same causal factors affect both A(t) and B(t), it is unlikely that the values of the causal factors will be the same in both cases. In using B(t) as an estimate of A(t) we can take note of the differences in the values of each of the causal factors and make adjustments in our predictions to take these differences into account. Although checklists of causal factors can be provided, the method requires experts to use their experience in identifying the most important causal factors to use.

The general CBP strategy (outlined in Table 1) begins with the definition of the target variable, A(t) and the identification of major determining ("causal") factors known to affect it. Next, a selection of possible comparison cases are identified.

From these, subject matter experts choose one case, based on the similarity—between it and the target case--of the effect of the causal factors. The comparison case variable that is analogous to the target case variable A (t) is specified as B(t). Subject matter experts then make a rough estimate of the differences expected between B(t) and A (t), often only a judgement of whether A(t) will be greater or less than B(t). They then are guided through an examination of the effect of the causal factors on the expected differences until this effect can be quantified so as to produce an applicable adjustment factor figure. This factor is then applied to operational data for B(t), to yield a prediction of A(t). Analysis of the differences among factors produced by SMEs can produce a confidence range for their prediction.

This CBP technique relies on the use of Subject Matter Experts (SMEs) who are knowledgeable about the domain of interest, in order to select optimal comparison cases and identify the relevant causal factors. The CBP approach elicits SME judgements through the use of a carefully structured interview with a format reflecting the CBP process outlined in Table 1. The approach is data driven since the SMEs are generating ad-

Figure 2

Logic of Comparison-Based Prediction

PREDICTION TARGET A(T)

CASE A:	(X, Y, Z, ...) → A(T)
COMPARISON CASE B:	(X, Y, Z, ...) → B(T)

Table 1

CBP Process

1. Identify the Prediction Target: A(t).

2. Prepare a checklist of major determinants of A(t) -- "Causal Factors" (x, y, z . . .)

3. Identify potential Comparison Cases.

4. Select the best Comparison Case(s) on the basis of similarity of Causal Factors.

5. Obtain data for B(t).

6. Estimate whether A(t) is expected to be greater or less than B(t). Use the Causal Factors to explain the difference between A(t) and B(t).

7. Adjust the B(t) data to generate a data prediction for A(t).

8. Estimate prediction boundaries as a confidence range.

justment of operational data, and giving their reasons for making these adjustments. There may be cases where no operational data are available. It is possible to proceed with a CBP approach by having the SMEs estimate the operational data as well as the adjustment factor, but this is not the ideal application of CBP method, and will reduce confidence in the outputs. Unfortunately, this is usually the state of affairs for predictions about the effectiveness of training devices, where reliable operational data are rare.

Our analyses have suggested that there is an entire set of strategies for using comparison cases to generate predictions. The Comparability Analysis method developed by the Air Force is one type of Comparison-Based Prediction method, a relatively straight-forward one. The strategies may include the use of multiple comparison cases and multiple Subject Matter Experts. One potential problem with using a single comparison is that if you have selected the wrong comparison case, you will be mislead. The use of several comparison cases, if well chosen, produces multiple predictions that will converge in a more reliable prediction. A divergent prediction value may signal a poor comparison choice.

Since CBP relies on subjective judgement it is important to consider the predictive validity of the approach. The CBP approach has been used successfully in the logistics domain for making predictions about new situations. We have examined logistics data (Klein & Gordon, 1984) and found that correlations between predicted and observed reliability of subsystems on A-10 aircraft ranged from .36 to .84. The higher correlations were found for the cases where empirical data were available, and did not have to be estimated. To continue the evaluation of the predictive validity of CBP, as part of ARI contract MDA 903-83-C-0270, a separate study is being conducted to evaluate the accuracy of CBP predictions in a study of transfer of training being conducted at George Mason University.

There are also strategies that can be used to increase the reliability of the predictions. These include the use of several SMEs, and/or the use of several comparison cases, to derive predictions of the same A(t) to see if these predictions converge on the same value. In addition, the creation of an audit trail can help a user understand how the prediction was generated, and to see the causal factors that went into the prediction.

The use of CBP has a number of advantages over other prediction techniques. It is useful when some parameters are unknown, data are missing, or objectives are unclear. Formal models can appear to increase predictive validity, but are difficult and time consuming to construct. It may not even be possible to develop them unless we have a great deal of knowledge about all of the relevant causal variables. CBP is a way of estimating unknown properties of a new situation on the basis of existing data that are modified using carefully elicited expert judgements. It requires relative judgements which seem to be easier to generate than absolute judgements about predicted values. The approach uses the concrete experience that already exists in most situations. It creates an audit trail of the prediction process which allows evaluation and re-adjustments, if necessary, to refine the prediction. Finally, it has face validity since it is an extension of the natural reasoning process of analogical inference. (Table 2).

The Use of CBP in Predicing Training Effectiveness of Prototype Maintenance Training Devices (TDs)

The goals of the project described in this report were to (1) determine whether the CBP strategy could be used to predict the effectiveness of maintenance TDs (2) document its strengths and weaknesses; and (3) support the Army Maintenance Training and Evaluation Simulation System (AMTESS) program, by predicting the effectiveness of prototype maintenance training devices and by generating data for design recommendations from information collected using the CBP approach.

Rationale for using Comparison-Based Prediction.

A research program, SIMTRAIN, was initiated in 1981, sponsored by the Army Research Institute to provide analytically and empirically based guidance for the development of training devices and simulators. Honeywell, Inc. was awarded a contract to pursue this program, and Klein Associates was subcontracted to evaluate the use of CBP methods for evaluating the effectiveness of new maintenance training devices. The study was limited to the troubleshooting of the engine starting system task as taught in several courses at APG and Edgewood Arsenal. At that time, the courses were using actual equipment and a two-dimensional flat board trainer, the OMNIDATA device, to train students on the troubleshooting task. For the CBP analysis, the Grumman and S/B devices were evaluated using the OMNIDATA and the actual equipment as comparison cases. The CBP method was able to generate global predictions of hours saved by the prototype TDs, and predict Training Effectiveness Ratios (Klein, 1982).

The results of this initial and preliminary effort suggested that it might be feasible to use CBP for predictions about a range of training tasks. That is, the use of comparison cases was considered for the first time to have potential for predicting training device effectiveness. This report describes the attempt to use CBP to predict the effectiveness of TDs, across tasks, and for different courses. Because the two prototype training devices address different tasks, most of which are not taught on other training devices that could be used as comparison cases, the effort provided an opportunity to determine the applicability of the CBP strategy when the availability of comparison cases is poor.

Table 2

Advantages of Comparison-Based Prediction

* FLEXIBLE. Can be used with missing or ambiguous data for ill-defined objectives.

* STRUCTURE FOR EXPERT JUDGEMENT. The approach relies on experts, but it structures their inputs to increase validity and reliability. It asks for relative judgements, which are made with more confidence than absolute predictions.

* EXPERIENCE-BASED PREDICTION. The predictions are derived from operational experience, not abstract formulae. They reflect contextual variables that are otherwise difficult to include.

* AUDIT TRAIL. The prediction structure shows which comparison cases were used, how they were adjusted, and why. A manager can understand and evaluate the results, and even develop them further.

* FACE VALIDITY. The use of Comparison Cases is a common form of natural inference. Comparison-Based Prediction is a way to increase power and validity and also to make the prediction process more explicit.

The Target Cases.

During the AMTESS program, two contractors, Grumman Aerospace Corporation and a consortium of Seville Research Corporation and Burtek, Inc., were chosen to develop prototypes of generic maintenance trainers. Both were designed to address maintenance courses taught to Army personnel at Aberdeen Proving Ground (APG) and Edgewood Arsenal, in Maryland. The Grumman device is used to simulate the starting and charging systems for the M110A2 (Self Propelled Howitzer). The Seville/Burtek (S/B) simulator was designed to be used in training maintenance of a diesel engine. In addition, by changing some of the hardware and software components, the simulators can also be used to train missile systems.

These were the two training devices examined in the present study. (In generating some of the conceptual design recommendations, we also used comparisons to versions of these two TDs that had been modified to handle missile training rather than automotive training).

OBJECTIVES OF THE STUDY

The primary objectives were to determine whether CBP could be used to estimate the effectiveness of prototype training devices, and to learn how CBP methods might be improved. These serve a longer range goal of developing an easy to use guidebook on how to apply CBP.

In addition, there were several subordinate objectives in support of the AMTESS program. Three factors were important in determing them: The prototype TDs addressed different tasks and could not be compared to each other; the AMTESS effort was pioneering, so there was a limited availability of comparison Army maintenance TDs; and there was a concensus among SMEs and contract monitors that the TDs would not be introduced as a replacement for Actual Equipment Trainers (AETs), but would augment the use of AETs. Because of these factors, the following objectives were developed:

1. Predict the training effectiveness of the Grumman prototype device when both the TD and an AET are used together in one course. The comparison case was the AET as it was currently used in the course, without the TD.

2. Predict the training effectiveness of the S/B prototype device when both the TD and an AET are used together in the same course. Again, the comparison case was the AET alone.

3. Predict the training effectiveness of the Grumman prototype TD using TDs from another domain as comparison cases.

4. Predict the training effectiveness of the S/B prototype TD using TDs from another domain as comparison cases.

5. Generate conceptual design recommendations. (A.) A direct feature by feature comparison between the two prototype devices and (B.) a comparison between each of the TDs and the AET and maintenance TDs from another domain.

METHOD

Training Device Characteristics

The Grumman device. The Grumman TD was developed for training tasks involved in the troubleshooting (T/S) of the starting and charging system of a self-propelled howitzer. It was designed to help students learn to use the technical manual and STE/ICE (Standard Test Equipment/Internal Combustion Engine).

The configuration of the Grumman device (Table 3) consists of three major components. The 3-D simulator is a system which looks and acts like the M110 howitzer except that it doesn't actually operate. (As noted above, the 3-D component is interchangeable. Another 3-D component, not the focus of this study, was developed to train missile maintenance tasks). The student station is composed of a color TV, WICAT computer, and videodisc. The computer monitors the work done on the 3-D module and gives feedback to the student on the student station. The TV is equipped with a touch panel which allows students to enter responses directly on the screen. The instructor station consists of a keyboard, CRT, and printer (all connected to the computer system). The instructor station provides performance information in the form of time and errors per task.

The Seville/Burtek device. The S/B training device was designed to have interchangeable component so that a variety of courses could be taught using it (Table 3). The TD consists of a "core" set of components plus a simulator for automotive tasks or a simulator for missile tasks (the automotive tasks were the dominant focus of this study). The core components consist of the student station (CRT, display unit, and push-button response panel), a hardcopy printer, and the instructor station (CRT, keyboard, and control panel). The components are driven by a computer and Winchester disc system. The automotive hardware (i.e., 3-D component) consists of a full-scale diesel engine model (cylinder block with accessories) and instrument panel, engine controls, removable components, adjustments, sensors, and STE/ICE.

Courses and Tasks

The Grumman TD was evaluated in the context of the 63D30 course (Self Propelled Field Artillery Supervisor). The 63D30 course is specifically geared for the M109 and M110 howitzers. The students are all Non-Commissioned Officers (NCOs) with considerable experience and familiarity with the howitzer. The goal of this course is to train students as

Table 3

Comparative Matrix of AMTESS Design Characteristics

		GRUMMAN	SEVILLE
1.	Principle of Simulation	Simulated, generic	Simulated, actual(models)
2.	General configuration	Independent student control station	Satellite, instructor mediated
3.	Instructor station	Same as student station	Separate instructor station
4.	Student stations	Single-student	Single-Student
5.	Graphic display	Color CRT video disc 17" tube	Slide Projector Screen & CRT
6.	Alphanumeric	BW CRT 12"	CRT BW
7.	Input device	Touch panel (finger)	Function keys on response panel
8.	Keyboard	Required for instructor only	Required of instructor only
9.	Test equipment	Actual measurements simulated T.E.	Actual measurements simulated T.E.
10.	3D Hardware	Family generic, context* preserved	Family, models, context* preserved
11.	Operating System	Pascal, fixed RAM, convertible to ADA	RT-11 DEC
12.	Program storage	Floppy disc	Winchester Disc w/floppy backup
13.	Program language	Pascal (convertible to ADA)	ARIC (Burtek proprietary)
14.	Authoring system	Instructor-easily used	Instructor, medium difficult
15.	Portability	High	Medium
16.	Interfaces	2D-3D, selectable to 110K baud printer	Printer RS-232 9600 baud
17.	Modularity	Good	Medium
18.	Diagnostics	To individual module	Test program, diagnostics fo computer
19.	Repair procedure	Substitution of modules	None specified-from T.M.
20.	Motion	Full-plus freeze	None
21.	Audio	Full	None on radar, engine sounds on diesel
22.	Fault insertion	Simulated, program mediated	Simulated, instructor mediated
23.	Motor skills training	Via 3D simulation hardware	Model of actual equipment
24.	Actual equipment	Not required but usable	Not required not usable

*Context is that of a diesel powerplant

mechanics who will work with vehicles in field and shop settings. They are taught both to troubleshoot (T/S) and to Remove and Replace (R/R) faulty components. On the other hand, the 63H30 course is for training maintenance on a variety of vehicles. There is less familiarity with the howitzer and emphasis is on R/R more than on T/S. The Grumman prototype itself places greater emphasis on T/S tasks and less emphasis on R/R tasks.

The Seville/Burtek TD was evaluated for the 63H10 course (Direct Support Vehicle Repairman) and the 63H30 course (Organizational Maintenance Supervisor). The 63H10 is a Wheeled Vehicle Repair course which includes T/S and R/R tasks for a diesel engine. Students in this course have completed basic training but have had little experience with Army equipment. The 63B30 students are NCOs who have previously been trained as organizational mechanics (and have had some experience with the diesel engine). The course includes both T/S and R/R tasks. The S/B prototype places stronger emphasis on R/R than on T/S, compared to the Grumman Device.

The predictions of training effectiveness were obtained for individual tasks for the various courses. The task chosen for this purpose are listed in Table 4.

CBP Strategy

The basic goal of this application was to (1) guide SMEs in generating predictions of training effectiveness for the two prototype devices, and (2) derive design recommendations for future AMTESS training devices. In order to accomplish these tasks, we had to determine the target variable to be predicted A(t), the task(s) trained, the comparison case(s) which would be used to help generate the estimates for the prototype devices, and the set of causal factors which were perceived to have a possible influence on the target variable.

Identification of Target Variable: Training Effectiveness. "Training effectiveness" is an abstract concept which reflects the ability of the TD to aid in the instruction of the student. An effective TD may provide a number of benefits, such as a higher level of student expertise (e.g., knowledge of correct procedure), fewer errors, a shorter time needed for training, and faster performance on the actual equipment. The target A(t) selected for this application was the number of hours saved by using a TD for training a particular task. The selection of this variable had several advantages: it could be readily understood by the SMEs who would be generating the estimates; it could be useful for validating the CBP procedure (an empirical validation was not planned as part of this study because that would have required the procurement and use of the training devices in the actual courses); and the prediction of hours saved could be useful to personnel making decisions about training device procurement. Because training time differed for different tasks, courses,

Table 4

Course Tasks Used in Predicting Training Effectiveness
(Task descriptions are derived from training materials)

<u>Grumman</u>

63D30 and 63H30

Task 1: Perform tests on electrical system (continuity test with STE/ICE, resistance test with STE/ICE, DC voltage test with STE/ICE, AC voltage test with STE/ICE

Task 2: Troubleshoot electrical system (starting system, generating system, battery power system)

<u>Seville/Burtek</u>

63W10

Task 1: Troubleshoot engine starting system

Task 2: Troubleshoot oil pump failure (organizational and direct support)
- Perform organizational troubleshooting
- Perform direct support troubleshooting
- R/R oil pump filter and oil pump

63B30

Task 1: Troubleshoot oil pump failure (organizational T/S only)

Task 2: Adjust alternator drive belt

Task 3: R/R starter motor

Task 4: Inspect electrical system

and instructors, this variable was converted to "% of time saved" to aid in data synthesis and comparison. To summarize, CBP was used to generate predictions of the training time that would be saved by using the TD with the AET as opposed to using the AET by itself.

Comparison Cases. Two converging approaches were taken for the current research project. (Figure 3).

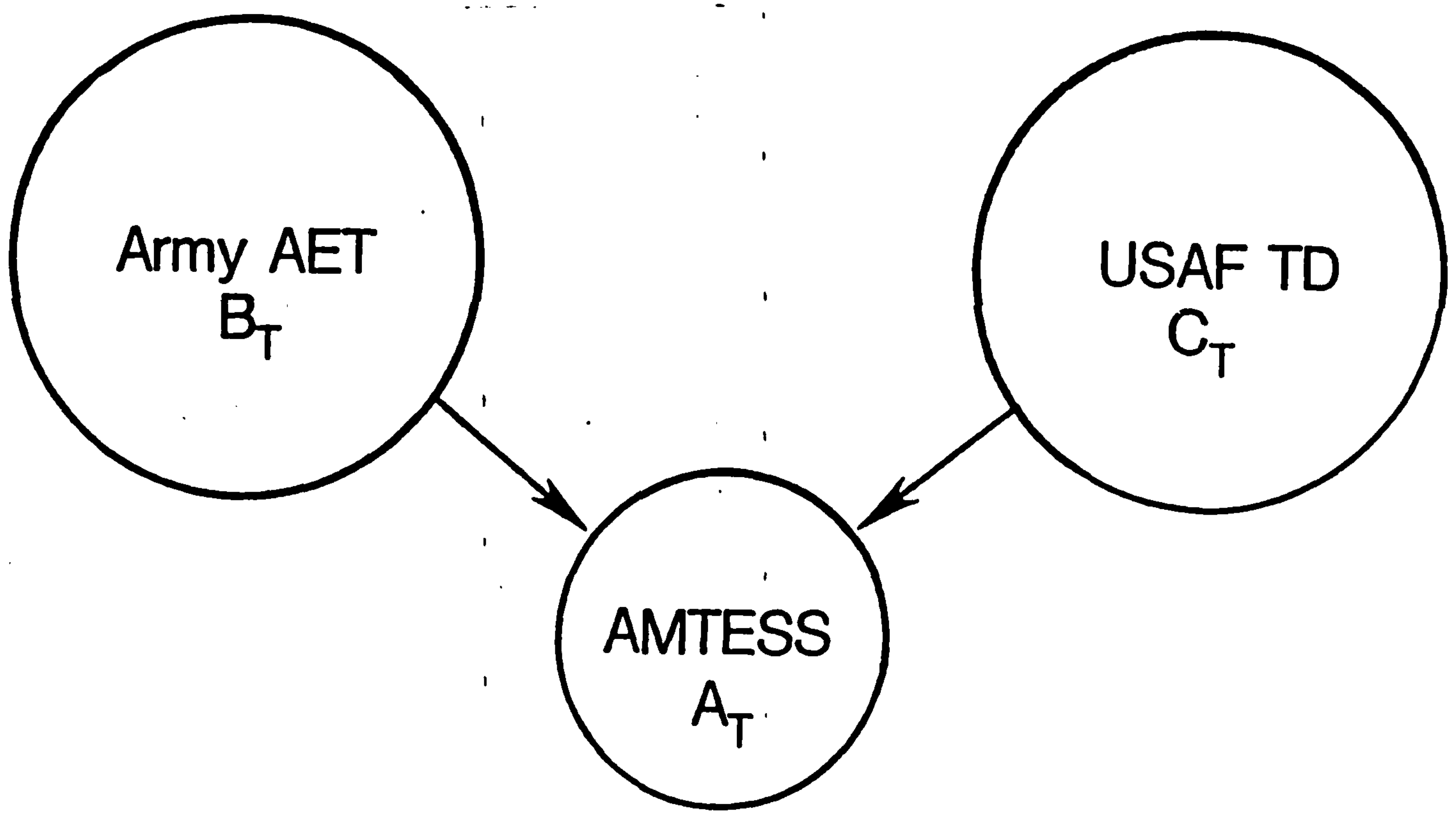
Since the course instructors were using operational equipment (AETs) for hands-on training, the AET was chosen as a comparison case. The SMEs were therefore asked to use their experience with training on the AETs to generate estimates of training hours saved by using the prototype device along with the AET.

An important problem was the lack of data on training effectiveness for the comparison training device (in this case the AET). To resolve this problem, the SMEs were asked to estimate B (t), as the number of hours that it took them to train a particular task on the AET. This number was then adjusted to estimate A(t), the time it would take to train using the prototype device together with the AET.

A second approach was to make predictions using additional comparison cases. Ideally, these would be other maintenance TDs. Even if these TDs were not designed to train automotive maintenance tasks, they could serve as effective contrast to the AET, which was not a computer-driven dedicated instructional TD. Two appropriate devices were located at Lowry AFB; the devices were designed to train maintenance personnel to use the 6883 AIS (Automated Intermediate Station), a very complex and expensive diagnostic testing station used for troubleshooting flight control systems of the F-111 aircraft. The actual test equipment was the 6883 AIS: the AET is a large diagnostic unit used to troubleshoot electronics systems in the F-111 aircraft. When a malfunction is suspected in the aircraft, Line Replaceable Units (LRUs) are removed from the aircraft and attached by cables to the AET. By simulating the operation of the aircraft, the AET permits troubleshooting in the workshop setting. A complex patch panel allows personnel to pinpoint problems in the electrical flow between units.

There were two TDs designed to teach AF maintenance personnel to use the 6883 AIS. One was a Honeywell 3-D Simulator: Components of the 3-D simulator consist of (a) two computers, one of which drives the other, (b) an instructor console including the capability to monitor student performance and input programs, (c) a student console with slide projector and screen, and a keyboard for the student to enter his or her responses to questions displayed on the CRT, (d) a 3-D simulator of the AET unit designed to appear and operate as the 6883 AET, (e) mock-up LRUs with simulated cables (no patch panel was simulated due to high cost), and (f) a printer for recording student performance.

Figure 3



Schematic diagram of the use of multiple comparisons. A(t) is the predicted effectiveness for the AMTESS prototypes. B(t) is the estimated data for the AET. C(t) is the effectiveness data for Air Force maintenance TDs.

The second TD for the 6883 AIS was the The Burtek 2-D Simulator: Components consist of (a) a master console including a keyboard and slide screen, (b) a student action panel which allows the student to identify the action to be taken on a defective component, (c) a printer to give student performance data, (d) flat panel mock-up of the 6883 AET, (e) a panel which simulates the action of the LRUs , and (f) three part task trainers which are used to provide training on an oscilloscope, a patch panel, and the logic of electronic systems.

There were several reasons for selecting the 6883 Automated Intermediate Test Station TDs as comparison cases. First, they represented examples of maintenance training devices. By using them we could study the degree of convergence between predictions from two different sources: Army AETs and AF 6883 TDs. Second, empirical studies had been performed of the effectiveness of the 6883 TDs (Cicchinelli, et al., 1980; 1982) so that CBP estimates could be generated from empirical data on maintenance training device effectiveness. This follows from the suggestion that wherever possible, CBP estimates consist of adjustments to empirical data rather than to estimated data.

There were however, some limitations to the use of the AF 6883 TDs as comparison cases. The SMEs for example, who were interviewed were not Army personnel and were not familiar with the AMTESS prototypes, or with Army automotive maintenance tasks.

<u>Causal Factors</u>

After the determination of the appropriate comparison cases the next task was to identify the major factors which might casue the target variable (% of time saved) to differ from one training device to another. Our goal was to prepare a list of causal factors that we could present to the SMEs as a basis for making the necessary adjustments to data from the comparison cases, to produce an estimate of the training effectiveness of the prototype maintenance TDs.

<u>INTERVIEWS</u> We interviewed several automotive maintenance instructors in Dayton to identify causal factors that might affect course time needed to train specific tasks. These interviews were used to modify the causal factors initially identified by Klein (1982). The final set consisted of four causal factors.

(1) Training Potential: This concept reflected the instructors' judgement of the ability of the TD to adequately train the task of interest. Thus a device with realistic 3-D hardware and carefully thought out instructional software has high training potential, whereas a device with unrealistic 3-D hardware and inadequate software has low training potential. This factor addresses the instructional value of the device, as opposed to the logistics of using the device in a course.

(2) Utilization: This factor emphasized the efficiency of running students and managing the logistics of training. For example, a TD whose design requires students to stand around for long periods of time while the instructor sets up malfunctions may have reduced utilization compared to a device that has little "dead time".

(3) Availability: This factor referred to the amount of time the training device was available to the instructor. An unreliable device that is frequently not running or is producing errors is not available for training. A device that is generally reliable but takes a long time to repair is also low in availability. This causes major training problems for the instructor. This causal factor was an important one to use in this application because the poor availability of one of the AMTESS devices was an important consideration in its effectiveness. It is true that prototype devices cannot be expected to have the same availability as operational devices. However, by specifying availability as a separate causal factor, we were in the position of being able to factor this out of estimates of course time saved.

(4) Motivation: This factor referred to the degree to which instructors judge that device characteristics motivate the students to learn the task. For example, 3-D simulation may motivate the student, whereas a long delay between tasks may decrease student motivation. Conceptually, motivation can be distinguished from training potential; it can be viewed as a subset of training potential.

This set of causal factors was used as a basic checklist in structuring the SME predictions. However, SMEs sometimes elaborated during the interview sessions (see below) and described other causal factors which they felt were important.

Subjects

Subjects in this project were 10 SMEs chosen on the basis of their familiarity with either the prototype devices or the comparison devices. The subjects are described in Table 5, and will be referred to by number in the remainder of the report.

The subjects generally fell into one of three categories. SMEs 1-5 were instructors who have taught the maintenance courses using only the operational equipment as training devices or using only the prototype devices (for evaluation research). Their experience ranged from 1 1/2 to 3 years.

SMEs 6-8 were professionals familiar with the Air Force 6883 AIS training devices (the Honeywell 3-D simulator and the Burtek 2-D simulator).

Table 5

Summary of SME Background

Maintenance Course Instructors	Location	Area of Expertise
SME 1	APG	Course 63D30
SME 2	APG	Course 63H30
SME 3	Edgewood Arsenal	Course 63W10
SME 4	Edgewood Arsenal	Course 63B30
SME 5	Edgewood Arsenal	Course 63B30
Air Force 6883 AIS		
SME 6	Air Force Human Resources Laboratory	Evaluation of Honeywell 3-D and Burtek 2-D simulators
SME 7	Honeywell	Design Engineer
SME 8	Honeywell	Research Psychologist
Prototype TD Evaluators		
SME 9	APG	Evaluated TD for automotive tasks
SME 10	Ft. Bliss	Evaluated TD for missile tasks

Finally, SMEs 9 and 10 were employees of Science Applications, Inc. who have been involved for the past year in a contract evaluating the two prototype devices. Both were thoroughly familiar with the configuration and capabilities of the devices. Neither were experienced with training a task within the existing courses. Their data were used in comparing the specific features of the two prototype devices.

The selection of SMEs is a difficult problem for any technique depending on expert judgements. Who is an expert? How do you verify expertise?

The primary type of SME used in this study was the course instructor. instructors qualified as SMEs because they were directly involved in training the tasks of interest. However, they are generally unfamiliar with psychological concepts in instructional theory, and may have difficulty in using analytical methods for front end analyses. Therefore, by using course instructors, we could also test the range of applicability of CBP: if CBP were too complex a procedure to be used with course instructors, if it required judgements that course instructors were not able to make, then its range of applicability would be severely limited. For each course, there was clear agreement about the one course instructor with primary knowledge, and there was never more than one course instructor per course so these were the four SMEs used. A fifth SME was run, a course instructor who had used the S/B prototype to some extent, but he quickly indicated that his knowledge was much less than that of the major SME identified, and his data were used only as a source of comments and questions.

For knowledge of the comparison TDs used by the Air Force, there were a number of SMEs who might have been chosen. The three selected included a contract monitor for the evaluation study and two contractors involved in device preparation and implementation. The primary criterion was availability of SMEs. There were no other reasons to value the inputs of one SME over those of another.

Regarding the two SMEs from SAI, these were the two individuals who had worked directly in the evaluation of the AMTESS prototypes. There were no other individuals with greater familiarity with the prototypes.

Materials

For each SME, materials were needed to describe the task, to explain how to do the comparison, and to collect the data from the comparison.

A task list was derived, identifying each of the tasks of interest for each of the courses studied (Table 4).

A data collection form (Appendix A) was prepared, to serve as a checklist for information to be collected from each SME. The form also included the checklist of causal factors that had been identified through preliminary research; it was useful to show this to SMEs to help them derive their estimates.

Documentation was prepared describing the prototype TDs. This was needed for the SMEs at APG and Edgewood Arsenal, because they had not worked with the TD in approximately one year. It was even more necessary for the SMEs who had worked on the Air Force 6883 training program and were totally unfamiliar with the prototype maintenance TDs. This documentation consisted of photographs of the two devices, feature lists, and a schematic drawing. Listings of the maintenance tasks that each device could be used to train were also made available.

Procedure

The primary data collection activity was the one-hour interview during which CBP information was obtained. We spent ten minutes explaining the purpose of the study, and gathering background information about SME experience, course duration, and other characteristics.

We spent approximately 15-20 minutes collecting data for each task studied for each of the two TDs being evaluated. After each interview, we added up the estimates made for each causal factor and derived an overall estimate of time saved/expended. This total was presented to the SME.

Follow-up calls to SMEs 1 through 9 were needed to verify data and to reconcile differences between SMEs as well as differences between the data and narrative for an SME. An advantage of CBP is that it tends to make explicit the prediction assumptions. It does this by analyzing differences as the result of differences in causal factors. For example, this enabled us to question one SME about the estimate for availability if data on this factor from another SME were nor consistent with his. We could then find out if there was in fact a disagreement, or if the two SMEs were using the concept of availability of mean different things.

RESULTS: PREDICTIONS OF TD EFFECT ON COURSE TIME

This section presents the prediction results separately for the two prototype TDs. Following these analysis there is an evaluation of the issue of how instructor time is saved by the TDs. Implications for future training device design are discussed separately below.

Grumman Prototype For the Grumman device, two course instructors were interviewed. SMEs 1 and 2.

Table 6 presents the data for SME 1, the instructor for Course 63D30. The table shows the impact of the causal factors for training with the AET alone, vs. training with the TD and the AET used as a supplement. The data represent the number of minutes each task would require for training. For example, Task 1 would take approximately 105 minutes to train using the AET. If the same task were taught using the TD, because of differences in training potential and availability, that task would require 150 minutes on the TD and an additional 60 minutes supplemental training time if the TD was used in Task 1. There is also a 100% predicted increase for Task 2. Combining the two tasks, the increase for the course is obviously 100%. Most of this was a function of availability, as shown in Table 6. The TD was estimated to have an availability rate of only 45% to 50%.

These low availability rates clearly preclude the use of this prototype as a TD in the course. Of course, prototype are not expected to have the same availability as operational training devices. If availability were improved greatly, would this make the Grumman TD useful for training? Since we specified availability as a separate causal factor, we were able to factor it out of the prediction. The remaining impact would still be for a time increase of only 20% for the overall course.

Several reasons were given for the large amount of supplementary AET training included in this SME's estimates. This need indicates that the TD is not adequately training certain tasks. Therefore, these reasons are specific examples contributing to the causal factor of training potential. For example, SME 1 felt that the TD was geared down to a 10 level student, rather than the 30 level students he has in his course. Related to this problem, his course includes a number of students who were familiar with the STE/ICE equipment. These students did not learn very much from the TD training. SME 1 felt that their time was more effectively used when they just had the AET. The TD used STE/ICE only as a multimeter, whereas the STE/ICE actually can perform many additional functions. This suggests the need for better information on how to design TDs to meet the needs of specific types of students. Even worse, the TD did not do a good job of teaching students to use the Technical Manual. The TD teaches by the Technical Manual and tells the student the page. However, students never have to learn to find a page by themselves. When they get back on the floor with the actual task, they are not oriented to use the manual independently.

There were also problems with utilization. The TD included a lock-step approach that forced students to go at the pace of the TD. Although the TD had the capability to skip instructional steps, this took the instructor time to cut out the unnecessary segments and it disrupted the student performance feedback because the software was not designed to take into account changes in the lesson sequence. In sum the software made it so difficult to use this capability that SME 1 was reluctant to take advantage of it. SME 1 also complained that the slow computer re-

Table 6

Predicted Time Impact for Grumman TD
in Course 63D30, for SME 1

	Causal Factors	(1) AET Alone	(2) TD	(3) AET as Supplement	(4) TD Savings Minutes* %**
Task 1: Perform Electrical System Checks	Training Potential	105 min.	60 min.	60 min.	
	Utilization		-	-	
	Availability		90 min.	-	
	Motivation		-	-	
		105 min.	150 min.	60 min.	-105 min.-100%
Task 2: Trouble-shoot Electrical System	Training Potential	120 min.	60 min.	90 min.	
	Utilization		-	-	
	Availability		90 min.	-	
	Motivation		-	-	
		120 min.	150 min.	90 min.	-120 min.-100%
Combined Tasks:		225 min.	300 min.	150 min.	-225 min. -100%
Combined Tasks Without "Availability" factor		225 min.	120 min.	150 min.	- 45 min. - 20%

*Column (1) minus [(2) plus (3)] (Negative figures denote an increase in training time)

**Column (4) divided by (1)

sponse time slowed down the training. In addition, utilization was affected by limitations in the instructions given to the instructors. SME 1 was never told how to repeat a lesson for a student. He discovered this on his own by accident. Another problem that wasted time was the need to punch information into the keyboard so that students could not see the nature of the exercise malfunction. It was not practical to isolate the instructor's station so the instructor had to ask the students to leave the floor area before entering lesson information. It might take another 25 minutes to round the students up again after their "break".

There were additional issues related to utilization. SME 1 noted that students needed to learn how to use the TD, and this TD-specific learning will have no generalization to the rest of their work. It takes time and confers no benefits. This opinion underscores the need for better information on how to design TDs to meet the needs of specific types of students.

SME 1 also complained that he was never shown how to fit the TD into his course, never advised how to change his course to make better use of the device. He pointed out that although 10 instructors at APG learned to operate the Grumman device, most rejected the TD because it was unreliable or required too much learning from them, or because they were un comfortable with the computers. Although these issues relate to utilization, SME 1 treated utilization as a part of the causal factor of training potential.

He did not think there would be a large motivational effect. He judged that the students who were familiar with STE/ICE tended to be motivated, but those familiar with STE/ICE disliked the TD. These opinions underscore the need for an easy way to skip lessons.

It is worth noting that this SME considered himself a strong advocate for the maintenance TDs. He had become discouraged because of the poor availability, low level of instruction, and other inadequacies of implementation, but he still felt that a properly executed TD would have definite benefits, especially for training STE/ICE.

The course instructor for the 63H30 course, SME 2, presented estimates that are summarized in Table 7. The estimate is for an overall increase of 75.8% in course time if the TD was to be included with the AET for training. The increase is primarily due to the availability problems experienced with the Grumman TD. This SME estimated availability of 50%, compared to an availability rate of about 80% for the AET. Actually, he had several AETs to draw on, so that if one malfunctioned he could turn to another. If the poor availability was factored out, the remaining impact would represent a 17.7% savings if the TD was used. This is markedly different from the prediction for Course 63D30. The difference is probably due to the fact that the 63H30 students were not at all familiar with STE/ICE, and found the TD quite valuable. This is a reason why the train-

Table 7

Predicted Time Impact for Grumman TD in Course 63H30, for SME 2

	Causal Factors	(1) AET Alone	(2) TD	(3) AET as Supplement	(4) TD Savings Minutes*	%**
Task 1: Perform Electrical System Checks	Training Potential	180 min.	125 min.	17 min.		
	Utilization		-	-		
	Availability		240 min.	-		
	Motivation		-	-		
		180 min.	365 min.	17 min.	-202 min.	-112%
Task 2: Trouble-shoot Electrical System	Training Potential	205 min.	50 min.	125 min.		
	Utilization		-	-		
	Availability		120 min.	-		
	Motivation		-	-		
		205 min.	170 min.	125 min.	- 90 min.	-43.9%
Combined Tasks:		385 min.	535 min.	142 min.	-292 min.	-75.8%
Combined Tasks Without "Availability" factor		385 min.	175 min.	142 min.	68 min.	17.7%

*Column (1) minus [(2) plus (3)] (Negative figures denote an increase in training time)
**Column (4) divided by (1)

ing potential was higher in the 63H30 course. As further instances of training potential, the instructor appreciated the fact that students could not complete a module unless they knew how to do the task. With the AET, students making errors might try to get another student to complete the task for them. In addition, this instructor did not think that that the TD made it hard for his students to learn to find the right page in the Technical Manual.

The instructor for 63H30 also liked the fact that the TD standardized the instruction. With the AET, the training can vary with the instructor's motivation on a particular day. This is another example of training potential.

This course instructor felt that his students were initially motivated to use the TD, but this motivation disappeared because the TD did not let them self-pace. They had to sit through everything, regardless of whether they already knew it, and this lock-step approach was discouraging. Another problem was machine error. The machine would sometimes give an error indication even if the student had gotten the problem right. This was quite annoying to students.

Summarizing these data, the two SMEs agreed that the TD would increase the time needed for training by 75% - 100%. There were different estimates for the availability factor, but even when this factor is omitted the deficiencies of the TD created a problem. For the 63D30 course there would still be a decrement of 20%. Only for the 63H30 course, with students unfamiliar with the STE/ICE equipment, was there a projected savings of 18% if availability problems were eliminated by Grumman.

A second set of three SMEs 6, 7, 8, were interviewed to obtain estimates for the Grumman TD. These were individuals familiar with the Air Force's Honeywell 3-D and Burtek 2-D TDs procured by AFHRL/ID for use in training troubleshooting tasks for the 6883 Automated Intermediate Station. (Cicchinelli et al., 1980, 1982).

Data are available on these TDs for the amount of time needed to train tasks using the AET (the actual 6883 equipment) vs. the TDs. The task of these SMEs was to use those data as a baseline for predicting the AMTESS TD time savings, making the necessary adjustments in causal factor values that differed for the two situations.

Table 8 presents the combined task data and predicted time savings for the Grumman TD from these three SMEs. It can be seen that different causal factors are involved because the comparison is between the AMTESS prototype maintenance TDs and two other TDs with comparable features rather than a comparison with the AET. For this application, the significant causal factors were the difference in the nature and complexity of the task. The factors of motivation and availability were similar for both applications.

Table 8

Predicted Time Savings from Grumman TD in Courses 63D30 and 63H30
As Compared to Air Force 6883 TDs

	Causal Factors	Savings Minutes	Savings %
SME 6 (24 hour block of instruction)	Motivation	-144	-10
	Task Simplicity	240	17
		96	6.7%
SME 7 (16 hour block of instruction)	Faster Computer Response	60	6.2
	Task Simplicity	60	6.2
		120	12.4%
SME 8 (16 hour block of instruction)	Use of Videodisc	45	4.7
	Task Simplicity	60	6.3
	Availability	--	- -
		105	11.0%

For SME 6 the predictions were based on the comparison between the Army's Grumman TD and the Air Force's Honeywell 3-D TD and Burtek 2-D TD. One major causal factor was the difference in task complexity. The nature of the task was seen as much more difficult for the Air Force course, and this favored the Grumman TD since there should be a greater chance to simulated a simpler task. The impact of working with a simpler task was estimated as having a potential savings of 4 hours over a three day block of instruction. Availability was not included as an impact. If anything, this SME felt that availability favored AF TDs since they were much more reliable than the actual equipment, whereas the Grumman TD was much less reliable than the AET. Motivation was felt to also favor the two Air Force TDs. For the 6883 instruction, the AET was often not functioning or not available for hands-on training, so the opportunities for using the TD for hands-on training were much appreciated. However, for automotive maintenance training there is the reverse situation. Students are getting hands-on training, and the TD just duplicates this in a less realistic way, so that many students might feel cheated. The B(t) data were that the AF TDs took the same time as the AET. (This was considered a success, since they were less expensive). Applying this to AMTESS, task simplicity favored the Grumman TD, but motivation did not. The overall prediction was for a savings of 6.7% for the Grumman TD.

SME 7 made his comparisons between the Grumman TD and the Honeywell 3-D TD. The major factors differentiating the two were computer response and task simplicity. He felt that existing off-the-shelf computers were so much faster than the one in the Honeywell TD that a considerable amount of training time could be saved just by eliminating the dead time waiting for a computer response. He also felt that the Army maintenance tasks were simpler than the Air Force maintenance tasks (comparing a diesel engine to an Automated Intermediate Station for testing Line Replaceable Units (LRUs) from an F-111 aircraft) and therefore there should be a greater chance for an effective training device. For SME 7, videodisc offered a potential for time gains, but this was offset by lower availability as well as the concern he felt about the availability of the Grumman TD.

SME 8 responded most strongly to the use of videodisc by the Grumman TD, since he felt that this would allow a faster presentation of a wider variety of material, including motion displays of task performance. Availability was considered as a potential factor, but only if an off-the-shelf computer was used. If a state-of-the-art computer system was used, then there would be worse availability (as was also the case with the Honeywell 6883 TD). Task simplicity was estimated to lead to an hour of savings, over a 16 hour block of instruction.

To summarize, the results show that SMEs using separate sets of comparison cases agree in predicting a modest savings with the use of the AMTESS prototype TD. SME 6 predicted a savings of at least 6.7%, and indicated that the actual savings might be higher. The other two SMEs (7 and 8) predicted savings rates of 11% and 12.4%. The median of the three was 11%. These predictions are close to those for SME 2, who estimated a savings of almost 18% (17.7%) if the availability problems were somehow overcome.

We can compare predictions directly, because the target variable A(t) and scenario were the same for the SMEs; the differences among them were in the comparison cases (Bs) that each used. A composite prediction, assuming not availability problems , was for a savings in course time of 14.5% + 3.5% savings.

Although the predictions for SMEs 6, 7, and 8 were close to those of SME2, they did not agree with the estimate of SME 1 for Course 63D30 they did not anticipate that a TD would be developed in a way that rendered it unavailable 50% of the time, and focused on skills already obtained by the students. The composite prediction was for a 5% loss of time due to the TD, + 15%. This large variability suggests that low confidence should be placed in this prediction.

Seville/Burket TD, Course 63W10.

Table 9 presents the estimate generated from SME 3, who was the only course instructor used in the prediction of the time saved if the S/B TD was used in this course.

The overall estimate is for a savings of 16.7%. This is balanced equally between the two tasks: 22% savings for Task 1 and 22% savings estimated for Task 2 (with time needed to train students to operate the TD factored in on top of the estimates for each of the courses).

This instructor felt that the training device was valuable in his course. He felt that it was easy to use, and that his students were also motivated by it. In fact, this was major factor, along with availability. He estimated a total savings of 75 minutes, or almost 17% (16.7%), just because the students worked harder on the TD. He reported for task 1 that some students wanted to push him farther into the available lessons than he was prepared to teach them. For Task 2, the students were very pleased not to have to get messy with oil and grease and dirt. They worked harder and more steadily. When asked about the loss of training realism due to the cleaner conditions, he did not see that as a problem.

In terms of utilization, this SME liked the fact that he could easily insert malfunctions into the TD, whereas with the AET he had to send students out of the area when he wanted to bug some equipment. He liked the fact that the TD told him how students were torquing a bolt. With the AET he had to check performance by torquing each bolt himself after the student was finished. However, these factors were all combined under training potential.

Table 9

Predicted Savings for Seville/Burtek TD,
for Course 63W10, for SME 3

	Causal Factors	(1) AET Alone	(2) TD	(3) AET as Supplement	(4) TD Savings Minutes*	%**
Task 1: T/S Engine Starting Systems	Training Potential	240 min.	210 min.	60 ± 20 min.#		
	Utilization	-	-	-		
	Availability	30 ± 10 min.	-	-		
	Motivation		-60 min.	-		
		270 min.	150 min.	60 min.	60 min.	22%
Task 2: T/S Oil Pump Failure	Training Potential	180 ± 30 min.	180 ± 30 min.	-		
	Utilization		-	-		
	Availability		-	-		
	Motivation		-40 min.	-		
		180 min.	140 min.		40 min.	22%
	Training Device Learning ##		-25 min.		- 25 min.	
Combined Tasks:		7½ hrs. ± 40 min.	5 hr. 15 min. ± 40 min.	1 hr. ± 20 min.	75 min.	16.7%

*Column (1) minus [(2) plus (3)]
**Column (4) divided by (1)

#The estimate is for 60 minutes plus or minus 20 minutes, reflecting the fact that the SME felt that he would take as little as 40 minutes or as much as 80 minutes.

##This estimate was generated by SME 9, and used for SME 3. It has not yet been verified by SME 3.

He estimated that the S/B availability rate was 98%. The only problem he recalled was when the TD was left in an unsecured place and someone tampered with it, messing up the projector. In contrast, the availability of the AET was estimated at 80-85%, and this was a serious problem for Task 1. It was not a problem for Task 2 since he did not handle Task 2 as a troubleshooting task, but rather concentrated on having students learn to remove/replace the oil pump and filter.

Interviews were conducted with SMEs 6, 7, and 8, who were familiar with the Honeywell 3-D TD used by the Air Force. Their data are presented in Table 10. For SME 6, the estimated savings of the S/B TD was at least 6.7% and possibly higher if individualized sequencing could be imposed. SMEs 7 and 8 estimated savings of 12.5%.

Each of the three SMEs used his predictions for the Grumman TD (Table 8) as a basis for predicting A(t) on the Seville/Burtek TD. SME 8 modified his prediction in Table 8. He noted that the S/B device did not have videodisc; but it was relying on more proven computer equipment and he therefore predicted greater savings for it than he did for the Grumman TD. Clearly, these three SMEs were not sensitive to the differences between TDs and courses. Their predictions are serving as a baseline for these types of TDs for these types of course.

Summarizing Table 9 and 10, the predicted savings for the S/B TD for Course 63W10 was 15% ± 2% for the two tasks combined. The low end represents the predictions of SMEs 6, 7, and 8, and the upper end represents the prediction of SME 3.

Seville/Burtek TD, Course 63B30.

Two SMEs were interviewed, SME 4 and 5. The data for SME 5 were subsequently deleted. He explained that he was less knowledgeable than SME 4, and had not been formally instructed on the TD, but had been given some practice on it by SME 4.

The data for SME 4 are presented in Table 11. These data are not broken down into causal factors because he felt that there was no problem with availability of the TD, and there were only minor differences in utilization and motivation. These differences were included in the estimates of training potential. This SME felt that the major impact of using the TD came from differences in how the tasks were trained, how much supplementary training would be needed on the AET, and the time necessary to learn how to use the TD.

The overall estimate was that the S/B TD would produce a time savings of 9% if used to train four tasks studied in Course 63B30. These tasks were: 1) Troubleshooting oil pump failure (organizational troubleshooting only); 2) Adjust alternator drive belts; 3) Remove/Replace starter motor; and 4) Inspect electrical system. According to SME 4, the

Table 10

Predicted Time Savings for S/B TD (63W10 Course)
As Compared to Air Force 6883 Honeywell TD

Causal Factors	Savings Minutes	Savings %
SME 6 (24 hour block of instruction)		
Motivation	-144	-10
Task Simplicity	240	17
	96	6.7%
SME 7 (16 hour block of instruction)		
Faster Computer Response	60	6.25
Task Simplicity	60	6.25
	120	12.5%
SME 8 (16 hour block of instruction)		
Availability	60	6.25
Task Simplicity	60	6.25
	120	12.5%

Table 11

Predicted Savings For S/B TD, Course 63B30, for SME 4

	(1) AET Alone	(2) TD	(3) AET as Supplement	(4) TD Savings Minutes*	%**
Task 1: Trouble-shoot oil pump failure	48 min.	12.5 min.	10 min.	25.5 min.	
Task 2: Adjust Alternator Drive Belts	11 min.	9 min.	4 min.	-2 min.	
Task 3: R/R Starter Motor	60 min.	22.5 min.	12 min.	25.5 min.	
Task 4: Inspect Electrical System	40 min.	27.5 min.	10 min.	2.5 min.	
Training Device Learning		37.5 min.		-37.5 min.	
Combined	159 min.	109 min.	36 min.	14 min.	9%

* Column (1) minus [(2) plus (3)]
** Column (4) divided by (1)

greatest savings would be found for Tasks 1 and 3. For Task 1, he felt that the savings would result from the fact that the AET is very unreliable. The trucks used as AETs frequently have battery problems.

For Task 3, the removing and replacing of the starter motor is laborious and time-consuming. When performed with actual equipment, it takes two people to carefully maneuver the 50lb. motor. It is heavy and has poor accessibility and it is tricky to get in past the propellor shaft. The TD allowed the student to learn the task without going through this preliminary procedure, and the instructor felt that this was a prime area in which to save training time. He was questioned about whether the AET was teaching some useful tricks of the trade, but he insisted that any such tricks were irrelevant to the principles that his course was intended to teach. This suggests that the advantage of the S/B TD is to supplement AET when it is difficult or time-consuming to access the component. The student can learn more efficiently on the TD to diagnose, repair, and install; he can then learn how to maneuver with the AET.

For inspection and adjustment (Tasks 2 and 4) there is no advantage to using the TD. If there is a major training impediment, as in Tasks 1 and 3, then the TD can generate savings.

In general, SME 4 did not prefer to use the TD, even though he felt that the TD learning would be faster. He thought the TD was a good procedures trainer, but it was limited in the ways that students could make mistakes. There were fewer opportunities to put things on wrong, see where oil lines and gauges were connected, see how to fit tools onto adjusting blots in tight spots, learn how to remove panels to gain access to components.

The Honeywell 3-D TD was used as a second comparison case by SMEs 6, 7, and 8. These SMEs did not feel that there would be a substantial difference between the 63H10 course and the 63B30 course for the S/B device. Their estimates for the 63B30 course were therefore the same as those given for the 63W10 course (Table 10). Td gains between 6.7% and 12.4% were projected, with a median of 12.4%. Since SME 4 predicted an overall savings of 9% and the combined prediction for SMEs 6, 7, and 8 was 12.4%, the composite prediction is for a savings of approximately 11% ± 2%.

The general assessment of the S/B TD is that it does offer potential for saving training time, especially because it is very reliable. Its value is for training, troubleshooting (due to its efficiency in inserting malfunctions), and for training mechanical tasks that required difficult, time-consuming access to defective parts.

For Course 63W10, the course instructor felt there was a moderate savings using the efficiency of the TD to teach procedures, but its major value was motivational by creating a more pleasant and clean environment for learning. For the 63B30 course there was a potential gain that stemmed from the use of the TD to teach mechanical tasks that were awkward to perform using the AET. It should be pointed out that the efficiency of the TD was offset to some degree by the need for initial instruction on how to use the Td, as well as the need for supplementary AET time to show students where to find components on the actual equipment.

Instructor Time

TDs can save much time because instructors can quickly set up malfunctions and prepare training equipment (e.g., 30 seconds for a simple keyboard entry compared to one to two hours to bug actual equipment).

However, the SMEs did not anticipate large savings. Once actual equipment is bugged, it usually stays that way for all the students in the course, so the time expended per student is usually quite small. Instructors using the Grumman TD saw a continued outlay of time, since they had to spend so much energy keeping it running, making sure it was kept clean, maintaining temperature rates in the room, providing routine maintenance (tightening components, checking cables, etc).

For the S/B TD, the estimates for instructor time saved were higher. The instructor for the 63W10 course estimated a savings of one hour, based on the routine maintenance that had to be performed on the ACT, along with the need to repair broken components. SME 5 estimated a savings of 15 minutes for the TD just for saving the time needed to leave the area, walk over to the equipment room, get the tool box, and walk back. SME 4 did not anticipate any savings for the instructor using the S/B device. If there was a need to start up the engine prior to the course, he could let it run for 15 minutes while doing other jobs, so the only real savings was the 2 minutes it took to climb up to the cab and turn the engine on and off.

Therefore, the major potential for savings of instructor time was for tasks where there was some troubleshooting requiring the bugging of actual equipment and for tasks where the engine of the AET would have to be started, requiring maintenance of oil and battery levels. In contrast, purely mechanical tasks did not offer any advantages. In general, the savings to the instructor were not anticipated to be great.

RESULTS: CONCEPTUAL DESIGN RECOMMENDATIONS

Feature Comparison of Grumman vs. Seville/Burtek

Two SMEs were asked to make a direct comparison, feature by feature, of the Grumman and S/B TDs. The SMEs interviewed for the feature comparison were SME 9 for the automotive tasks, and SME 10 for the missile tasks at Ft. Bliss. The goal was to obtain estimates of relative effectiveness for the devices and to derive design recommendations from the comparisons. A data collection form was used for this section which consisted of a listing of causal factors for each of the feature comparisons.

Table 12 contains the feature comparisons obtained for both the automotive and missile tasks (where "+" indicates preference for either the Grumman or S/B device, and a blank space indicates that neither was especially preferred). The reasons for the preferences were included in the following section.

Composite Design Recommendations

The design recommendations are based on the interviews with all of the SMEs. The five course instructors compared the two prototype TDs with the AETs. The three AF 6883 personnel compared the prototype devices with the 6883 training devices at Lowry AFB. The two AMTESS evaluators compared the prototype devices directly with each other. In all cases, the proponents of a recommendation are indicated in parentheses at the end of the recommendation: SME 1-5 were the course instructors, SMEs 6-8 used the AF 6883 TD as a comparison case, and SMEs 9 and 10 were the SAI evaluators.

The intent of this part of the research project was to determine what types of recommendations could be made just by using comparison cases. That is, rather than starting the design phase from scratch, how much could we specify as modifications based on judgements obtained during the CBP process?

The intent of gathering these recommendations was to use CBP to identify design issues and to make some initial suggestions. Clearly, the data collected from these SMEs will not be directly useable by a design engineer preparing to write simulator specifications about the types of features that an operational simulator may include. The recommendations are a starting point for the training device design process, and would require additional interviews and data collection to develop the level of detail needed to include them in a set of specifications.

Summary of Design Recommendations

The recommendations fall into four categories: hardware; software/courseware; instructional capabilities and utilization; and ivailability. Our discussion of these recommendations assumes that an improved maintenance TD would be used in the same courses that were involved in the current evaluation (63D30, 63H30, 63W10, and 63B30).

HARDWARE 3-D Module - The Grumman approach was favored over the S/B approach (SME 9. 10). The instructions felt that the physical realism of the S/B device was not that necessary (1.4), and even with the more realistic S/B device the students needed supplemental AET orientation about

Table 12

Feature Comparison for Automotive & Missile Tasks

Feature	SME 9 Automotive Tasks		SME 10 Missile Tasks	
	Seville/Burtek	Grumman	S/B	G
3D Modules				
Training Potential		+	+	
Utilization		+		
Availability				
Motivation	+			
Student Performance Record				
Training Potential	+		+	
Utilization	+		+	
Availability	+			
Motivation	+		+	
Instructor CRT				
Training Potential	+		+	
Utilization	+		+	
Availability	+			
Motivation	+			
Student CRT				
Training Potential		+		
Utilization		+		
Availability	+			
Motivation		+		
Visual Display				
Training Potential		+		+
Utilization		+		
Availability				+
Motivation		+		
Student Panel				
Training Potential		+		
Utilization		+		+
Availability	+		+	
Motivation		+		+

Table 12 (Continued)

Feature	SME 9 Automotive Tasks Seville/Burtek	Grumman	SME 10 Missile Tasks S/B	G
Editing System				
Training Potential	+		+	
Utilization	+		+	
Availability	+			
Motivation	+		+	
Request Help				
Training Potential		+		
Utilization		+		
Availability				
Motivation	+			
Performance Feedback				
Training Potential		+		+
Utilization				
Availability				
Motivation		+		+

where to find things on the vehicle. If anything, less realism than the Grumman device would be acceptable for electrical tasks, along with some means of letting the student know where and how to orient to the actual equipment. There was a general feeling that funds poured into physical fidelity may be better spent on courseware (6, 7, 8). Videodisc, with its capability for portraying motion, was seen as valuable here.

For mechanical training, the S/B was not seen as realistic enough for adequate training (4). It was too easy to remove and replace items, whereas the AET had many more ways to put things on wrong, etc. It was suggested (4) that the entire front end of a truck be used, to show the problem with finding and working on parts.

Thus, there is a dichotomy concerning the nature of mechanical vs. electrical tasks. For mechanical task, even the high physical fidelity S/B TD was only useful for teaching procedures. Actual development of skills and recognition of context would take much greater physical fidelity. For electrical troubleshooting, the lower physical fidelity of the Grumman device was still more realistic than necessary.

Students Performance Record. The S/B format was favored here primarily because the Grumman device did not present performance feedback during the lesson (9, 10).

Performance Feedback. The Grumman device was preferred (9, 10). The SMEs felt that it gave more useful information and displayed "warmer" messages, which affected student attitudes.

Student CRT. The Grumman approach was preferred (9, 10) because all information was presented on the CRT and was therefore easier for students to use.

Student Panel. The Grumman touch panel was favored here (6,7,8,9, 10). The S/B response keys took longer to learn and operate. The major problems with the touch panel were availability problems (possibly stemming from the computer system) and error tendencies. Other approaches such as a mouse input device were not considered since neither prototype contained such features. This shows a limitation of CBP, and suggests that users of such recommendations must be prepared to examine features not represented in the comparison cases.

Instructor CRT. The S/B arrangement was preferred because of the way student performance data were presented during the lesson and because of the more reliable computer system (9,10).

Request Help. The Grumman approach was favored (9,10) because the student merely had to touch the screen if help was desired. With the S/B the instructor had to be called in to help the student.

Visual Display. The Grumman videodisc was preferred (6,9,10) because of more flexibility and better graphics, along with capability to show how a R/R action was performed. The primary caveats were the expense needed to modify it (i.e., purchase a new videodisc) and possibly greater reliability problems. There was some feeling (7) that video was better for mechanical than electrical T/S because of its dynamic properties and that CAI graphics might be more suited to electrical T/S.

Editing. The S/B approach was preferred (9,10) because it was easier to use by the course instructors.

Computer. The S/B approach was preferred here (1,2,3,7,8,9,10). No one had any support for the Grumman TD on this dimension. The Grumman computer was long recognized as a source of problems. There had been similar problems with the Honeywell 6883 training device, where new technology was not reliable, and the user wound up doing the troubleshooting for the manufacturer.

Longer cables were requested (2). The cables for the Grumman did not allow satisfactory separation between the 3-D components and the student and instructor stations.

Software and Courseware

It was suggested that the Grumman language, PASCAL, was preferrable to the S/B proprietary language because it would be easier to program in Pascal and to make changes (6).

There was a general request from the course instructors for much greater complexity in the tasks involving the use of STE/ICE. More complex problems, more complicated cables, confidence checks, and demonstration that STE/ICE is not just a multimeter were all requested. This parallels the experience with the 6883 (6,7,8) that the more complicated T/S problems would have been more valuable than physical realism.

In defense of the personnel who designed the prototypes, it should be pointed out that the requirement to include STE/ICE was added very late i the procurement cycle. It is not surprising therefore that the STE/ICE capability is limited, nor should this be interpreted as a criticism of the companies who built the prototypes.

It is important that students learn to apply the Technical Manual (1,2,9). The Grumman device did not require this; it led students through the problems and did not prepare them for independent use of the Technical Manual. This was also noted as a failing with the 6883 training devices.

The courseware needs to show students what components really look like, and where to find them on the actual equipment. The failure to provide this was a prime reason for the need for supplementary AET experience. (1,2,3,4).

The courseware should place more emphasis on conceptual learning, especially with regard to T/S (1,6,7,8,9). The TDs resembled procedures trainers, yet there was an untapped capability for training a variety of troubleshooting tasks. This was also observed for the Honeywell 3-D device (6,7,8).

Instructional Support and Utilization

Instructor training should be improved for future devices; the instructors should be given suggestions about how to fit the TDs into the courses (1,2). They should be shown how to repeat lessons, and take over more instructional responsibility.

The instructor must be able to have better access to the computer than was possible for the Grumman device (1). There was a complaint among instructors that to set up problems and malfunctions required them to ask students to leave the training area, wasting a large amount of time in recessing and then reassembling (1,2). The instructors wanted some easy codes to set up student problems so that students working nearby could not see what problem type or malfunction had been entered.

There was a general dissatisfaction with the lock-step nature of the Grumman prototype (1,2). It was recommended that instructors be given the capability to skip steps and delete materials for selected students (1). Moreover, these deletions must be able to be made easily (e.g., without turning the computer off), and without losing the student performance records and feedback.

In evaluating the Grumman device it was recommended (1) that the student performance sheet be improved. One suggestion was to break performance into more categories such as dangerous errors as opposed to regular errors.

For future TDs, instructors need a simple troubleshooting guide to basic, common, and frequent malfunctions that they can deal with easily (1). There appeared to be some regular and simple problems with the Grumman TD that the course instructor felt he could handle without having to wait for contract maintenance.

For future TDs, the instructor station could be more simplistic than the ones included in the prototype TDs (6). Simulators built by Research and Development teams may include performance monitoring capabilities that are needed for research even though they may not have training value. It may be that instructors are more effective when they are watching students, rather than watching instructor CRTs. With several student stations, there may be a bigger payoff for freeing instructors to walk around and observe than for tying instructors to the CRT. It may be useful to re-examine the need for an instructor station.

A recommendation was made that the delays of turning the computer off and then on again be avoided (1,2). For lunchtime, instructors would close down the system, and then after lunch it was a laborious process to begin again.

Availability and Reliability

Greater attention needs to be paid to the tendency of the Grumman prototype to signal student errors that were never made (1,2). This has severe motivational effects. It had also been a problem with the Honeywell 3-D TD.

The S/B device had problems with bolts breaking through standard wear and tear (4). This may have had some instructional benefits, but it limited the availability of the TD until the repairs could be made. The Grumman device had major computer problems. In addition there was some problem with the videodisc drawer jamming.

It is recommended that a new TD be required to maintain 95% availability. The Honeywell 3-D TDs were estimated at 85% availability (6,7,8) using early 1970's computer technology. The design of the AMTESS prototype TDs appeared to be less complex to these SMEs, who felt that they should have the potential for greater reliability. Availability for the S/B device was fairly high. The Grumman device availability, however, was estimated to be only about 45-55%, creating substantial training problems.

Alternative TD Recommendations

The comparison of AMTESS to the Air Force's 6883 training devices generates a general recommendation. There is an interesting analogy between the Air Force 6883 Automated Intermediate Station and the Army's STE/ICE equipment. Both of these are designed to perform automatically a variety of different types of tests. Both are complex for the personnel using them, and both have barriers to training. For the 6883, the barrier consists of its poor availability and the restrictions on entering malfunctions. For STE/ICE, there was a restricted use of the equipment for fear of damage to the STE/ICE kit itself. According to SME9, it was considered expensive by the instructors and treated carefully, limiting the chance for hands-on training. Thus, for both items of test equipment there exist clear training needs not being satisfied by Actual Equipment Trainers, a condition essential for acceptance of training devices.

The Air Force embarked on a program for acquiring training devices for the 6883 test station. Along the same lines, there could be value in developing a STE/ICE training device, dedicated to the training STE/ICE applications. It could be a low fidelity device useful in a variety of courses. That is, STE/ICE represents a piece of test equipment that could be easily simulated and widely used, reducing concerns about shifts in MOS requirements. It is true that both Grumman and the S/B devices teach STE/ICE functions, but only in a limited was for specific courses. What is being proposed is a generic STE/ICE training device.

The comparison between 6883 and STE/ICE reveals clear points of dissimilarity. For the 6883, the test was to troubleshoot and maintain the test station itself, whereas for STE/ICE the task is to use the test equipment to troubleshoot other equipment. Even more notable is the disparity in equipment cost. The 6883 actual equipment costs millions of dollars, whereas the STE/ICE kits cost approximately $3,000 each, so it may not be cost effective to develop a generic training device for STE/ICE.

SUMMARY, DISCUSSION, AND CONCLUSIONS

The major goal of this effort was to examine how the Comparison-Based Prediction method could be used to estimate the training effectiveness of the AMTESS prototype TD and how CBP could be improved. A complementary goal was to generate recommendations about design improvements. These goals were accomplished. We were able to elicit predictions from SMEs about the impact of the prototype training devices on maintenance training courses, and identified a number of conceptual design recommendations that could improve the effectiveness of the devices. Although the study did not include empirical validation of the data, the estimates from the SMEs converged closely for three of the four courses.

For the majority of the SMEs, there were no comparable training devices in use since AMTESS is a pioneering effort within the Army. Therefore predictions were obtained by making comparisons with Actual Equipment Trainers. This is a severe departure from the ideal conditions for Comparison-Based Prediction, as used for logistics predictions. Nevertheless, the method was effective for collecting prediction data, although we would not expect the quality of these predictions to be as high as if operational data were available for comparison cases. As more maintenance TDs are developed, this problem should diminish and validation studies easily designed.

TD Predictions and Recommendations

Two prototype maintenance training devices were examined, a Grumman device and a Seville/Burtek device. A summary of the prediction data for the two devices is presented in Table 13. The estimates given are for savings in training time obtained by using the device along with operational equipment (aets) as compared with using the AET alone.

In discussion of the data and underlying rationale given by SMEs, each device will first be treated separately.

Grumman TD. For the Grumman training device, there were no predicted savings in training time for the two courses examined. In fact, an increase of 75%-100% in training time was predicted. Since the CBP approach attempts to make explicit the causal factors behind the judgement, it was possible to determine that the major weakness of the Grumman device was poor availability. However, even if the availability problem were completely eliminated, the judgements were for a net increase of 20% in training time in one of the two courses, and a savings in training time of 18% in the other course. The remaining weaknesses of the device involved limitations in instructional software and the presentation of tasks at too low a level. It is concluded that the limitations of the Grumman device were not inherent in the existing technology (either computer technology or instructional technology) but were due to special weaknesses in the Grumman approach.

Recommendations for the Grumman training device were derived by comparing its capabilities to the existing Actual Equipment Trainer (AET) training device. The major recommendation was obviously for a more reliable and well-tested computer system. The remaining recommendations concerned the instructional software. It was felt that the presentation was at too low a level for many of the students, that the interaction with the Technical Manual was too artificial, and that the instructor control over lessons and sequencing was too limited. There was also need for showing the student more clearly the differences between training device and actual equipment performance, in order to facilitate transition back to the actual equipment. There is a potential value in reducing physical fidelity; the strength of the Grumman device was in being able to teach procedural types of troubleshooting tasks efficiently, and physical fidelity does not have to be high to accomplish this goal.

The Seville/Burtek training device was higher in physical fidelity than the Grumman device. It was predicted to have greater training potential, largely because of its greater reliability. For the two courses studied, the Seville/Burtek device was predicted to have a good potential for time savings. This breaks down as follows - predicted savings of 15% + 2% for course 63W10 and predicted savings of 11% + 2% for course 63B30. The value of the S/B TD was seen in terms of efficiently training remove/replace procedures and effectively working with students on proce-

Table 13

Summary of Predicted Savings in Training Time
for Grumman and Seville/Burtek TDs

Training Device	Comparison Case AET	Comparison Case 6883 AIS*	Summary Prediction and Confidence Range
Grumman:			
Course 63D30	-20%**	11%	-5%±15%
Course 63H30	18%**	11%	14.5%±3.5%
Seville/Burtek:			
Course 63W10	17%	12.5%	15%±2%
Course 63B30	9%	12.5%	11%±2%

*Each prediction given is the median for the three SMEs.

**Predictions are based on savings assuming that the problem of availability is eliminated. (Negative figures represent an increase in training time) dural types of troubleshooting tasks. It also had the potential for large time savings if used for tasks that were time consuming on the AET because of the inefficiencies of wrestling with automotive equipment. Its value in the 63W10 course was in quickly showing the basics of the tasks to low-level students who might miss the conceptual lessons during their struggles with unfamiliar mechanical equipment. For the higher level course, 63B30, it is not clear that the training device has much to offer except for tasks where mechanical problems make AET training inefficient.

Recommendations for the S/B device included the need to have better instructional software, to train a wider variety of task for the STE/ICE equipment and to represent the complexities of using STE/ICE.

Recommendations for Future Maintenance TDs

An attempt was made to develop design recommendations for a next generation of AMTESS training devices. The primary use of CBP was to obtain feature-by-feature-comparisons between the Grumman and Seville/Burtek devices.

There was strong agreement between the SMEs on the design of the next maintenance TD. For most of the task, especially troubleshooting tasks, the physical fidelity can be lower: more like the Grumman TD than the S/B TD, and possibly even lower than that. (It is assumed that supplementary AET training would be available to handle the sub-tasks that were not adequately presented in the TD). A greater use of computer graphics and videodisc capabilities might constitute an adequate replacement for much of the three-dimensional hardware. For some of the R/R tasks, the recommendation was for higher physical fidelity as a means of providing the full complexity and difficulty involved. There was a recommendation to use the Grumman approach also for the student panel, the student CRT, and performance feedback, as well as the instructor CRT and student performance monitoring. The S/B computer architecture was recommended, but with a standard computer language rather than the proprietary one used. The editing approach should rely on the Seville/Burtek format.

The interviews with SMEs also generated a number of additional recommendations for a next generation of maintenance TDs. Some SMEs felt that it may be useful to simplify the instructor station. Availability was a major issue, with the goal of 95% availability presented as a reasonable objective for an operational TD. Clearly a prototype device would have lower reliability rates. Utilization efficiencies were recommended to make it simpler for instructors to enter malfunctions without disrupting the classes. With regard to the instructional software, there were many recommendations. The primary ones were the need for more complex STE/ICE and troubleshooting problems, better description of the relation between the TD and the actual equipment, and greater flexibility for the student to skip or repeat lessons.

The Concept of Generic Maintenance Training Devices

An overall assessment of the AMTESS concept suggests some strengths and weaknesses of the generic maintenance training device concept. The generic TDs do have the potential to produce savings for tasks that are time consuming to perform using the actual equipment, and they can be an efficient way to teach procedures and troubleshooting. However, the training efficiency can be offset by the need to devote time to training students to use the training devices. This time has no long-term instructional benefits. Time is also needed to teach students how the actual equipment differs from the training device. For most tasks, these two training requirements can offset the training efficiency of the devices. Furthermore, few of the course instructors felt that the training device could serve as a replacement for the actual equipment. For their courses, they felt that it would be necessary to have the actual equipment for almost all tasks.

The concept of a generic TD offers a great efficiency in terms of design and procurement, but these gains must be balanced against specific course needs. The utilization of TDs is often a function of how difficult it is to train tasks using actual equipment. This can occur because fuel and munitions are not available, actual equipment is unreliable, etc.

In such circumstances there is often a ready acceptance of TDs over technology. However, this can develop into a preference for TDs over actual equipment regardless of whether the latter already adequately trains. For a number of the tasks studied, there did not appear to be a clear need for a TD. We suggest that, except for research purposes, training requirements rather than technology availability should determine the need for TD development. We recognize, however, that AMTESS is an evaluation of the concept of generic TDs, rather than an attempt to design specific TDs to address training needs.

Another difficulty of using maintenance TDs concerns the shifting of courses and MOS designations. An MOS shift occurred during the AMTESS program, leading to the delivery of a TD that had no course in which to be used. This is not a rare event. The same thing happened on the Air Force's 6883 program. We can assume that this possibility will continue and that devices designed for courses may become obsolete when those courses are no longer taught. This may suggest a stronger reliance on software and less reliance on simulator hardware for training.

It is worth nothing that this situation is not analogous to the pilot training tasks. Pilots have tasks to perform, and these remain relatively constant, whereas maintenance tasks are continually being reorganized as different MOS specialties are created or combined. This presents a major obstacle to the design of enduring maintenance TDs. There are some of the motivations to retain flexibility, by using generic TDs.

Yet flexibility in itself is not a great virtue. Devices actually have to train something. The procurement cycle for device modification must be significantly faster than the cycle for MOS rearrangement.

These comments should not be interpreted as criticizing the AMTESS program. The examination of new technological applications has potential value. However, the application of TD design concepts for pilot training may not have a simple extension to maintenance tasks and care is needed to recognize the special constraints of maintenance training requirements.

Assessment of Comparison-Based Prediction

This application of CBP has been valuable for introducing a variety of factors that must be considered in using CBP to predict training effectiveness. For future uses of CBP, it might be valuable to describe some of the lessons learned.

There was not an adequate description of the tasks that we were studying. We simply selected tasks from the lists presented by the device manufacturers; it was only in the field that we learned about the subleties of the tasks. The same task might be taught differently in two related courses, or course instructors might be willing to ignore part of a task trained on a training device, assuming that on-the-job training would take care of it. For analyses at the task level, there is a need for a reasonably explicit task description.

The causal factors needed clearer definition and explanation to SMEs. Training Potential was a major factor; different estimates of training potential often stemmed from differences in the way tasks were trained in specific courses. SMEs included factors here that properly belonged in Utilization or Availability.

Utilization primarily referred to the efficiency of the TD design for training students; this was not often used but it reflected a component distinct from training potential and it pinpointed specific inefficiencies that were useful as the basis for recommendations. However, the SMEs did not use this causal factor by name, so its value must be questioned.

The causal factor of Availability was an important one, especially for the Grumman TD. The use of this factor allowed us to study training potential of the TD more directly.

The causal factor of Motivation was not always used by the SMEs, but where it was applied it then tended to have a large impact. It was also an important dimension used by the SMEs using the Air Force 6883 training program as a comparison base. However, for a number of the

SMEs;, motivation was simply combined with training potential. Therefore, training potential became a global category distinguished only from availability, for the course instructors.

In addition to this original set of causal factors, we identified other components involved in training time: time needed to train students to use the TD, and time needed to show students the differences between the TD and the actual equipment. It seems likely that we are dealing with three components of training time, which together make up the overall course time. These three components reflect time spent training the student. Each of these components may or may not be affected by the causal factors. A matrix is presented in Table 14 demonstrating this concept. The expanded set of causal factors listed in Table 14 was derived from comments made by SMEs during the interviews. It would be valuable to develop a generic set of specific causal factors, but in this application most of the specific factors were idiosyncratic to particular device characteristics.

Other variables which might be affected by the causal factors are components of the instructor's time that is spent outside of student training. These components include learning about the TD and incorporating it into the Program of Instruction (POI), programming the various malfunctions, and repairing the TD (or AET). An example of how these components might be affected is shown in Table 15.

When Comparability Analysis was first applied to the task of predicting Air Force weapons systems reliability rates, such as Mean Time Between Failures, the task was simplified by decomposing an aircraft into subsystems. For the AMTESS application, we have learned that we could decompose the task of predicting training time impact into various time components. This helps to make the task more manageable and accurate. Matrices such as those presented in Tables 14 and 15 may be useful for future research.

The strategy for eliciting judgements was also examined. For the initial SMEs, we simply presented the standard causal checklist and later asked for additional factors. However, this procedure may limit the types of responses generated by subjects, since it forces them to use an unfamiliar system. For the later interviews, we found it effective to begin by asking the SMEs what they felt the critical differences were. We attempted to record these under the categories in the causal checklist. Then we presented the remaining causal factors, to see if the SME wished to include any more dimensions. This approach seemed to facilitate subjects' responding. It also increased the likelihood of learning about new causal factors. However, it did prevent us from using a standard set of causal factors for all SMEs.

Table 14

Potential Impact of Causal Factors on Training Time

Causal Factors	Training Time		
	Student learns to use TD	Train Task on TD	Supplemental Training on AET
Original Set:			
Training Potential		+	+
Utilization	+	+	+
Availability	+	+	
Motivation	+	+	+
Alternative Detailed Set:			
Training Potential			
TD provides accurate feedback		+	
TD teaches use of the Technical Manual		+	+
TD shows where components are on the AET		+	+
Utilization			
Ease of programming malfunctions		+	
TD allows skipping of steps		+	
Computer response time	+	+	
Availability			
Reliability	+	+	
Ease of maintenance		+	+
Time required for maintenance tasks		+	+
General			
Difficulty of learning to use TD	+		
Information provided by manufacturer	+	+	
Standardized program of instruction		+	
Use of videodisc		+	
Requires reading skills	+	+	

Table 15

Potential Impact of Causal Factors
on Instructor Time

Causal Factors	Instructor Time: Learn TD & Incorporate into POI	Instructor Time: Time to Program Malfunctions	Instructor Time: Repair TD or AET
Original Set:			
Training Potential			
Utilization		+	
Availability			+
Motivation			
Alternative Detailed Set:			
Training Potential - No Impact			
Utilization			
Ease of programming malfunctions		+	
TD Allows skipping of steps			
Computer response time			
Availability			
Reliability			+
Ease of maintenance			+
Time required for maintenance tasks			+
General			
Difficulty of learning to use TD	+		
Info. provided by manufacturer	+	+	+
Standardized program of instruction	+		
Use of videodisc	+		
Requires reading skills			

The use of a second comparison cases was a new type of Comparison-Based Prediction strategy, one that we had described but had not previously implemented. The concept was to use several comparison cases to arrange a convergence, the higher the confidence we have in the prediction.

We found that the SMEs familiar with the Air Force 6883 training program were able to generate predictions about the effectiveness of the Army maintenance training devices. These predictions were in close agreement with the predictions generated by the Army course instructors for the 63B30 course, and for the 63W10 course. These predictions did not reflect the poor reliability of the Grumman TD; otherwise they would have been close to the prediction for the 63H30 course. Furthermore, the Air Force based predictions did not reflect the redundant training provided by the Grumman TD for the 63D30 course.

These findings raise the question of what can be expected from CBP. It is a method for explicating what is presently known, in order to produce a prediction along with a clear statement of the assumptions behind that prediction. However, no predictive method can be sensitive to unlikely events. CBP and any other approach must assume a reasonable and standard implementation of hardware and instructional software.

One other effect of using additional comparison cases was the emergence of new causal factors. SMEs 6, 7, and 8 were sensitive to differences between electrical and mechanical tasks, and higher level vs. lower level students, but were less sensitive to differences within types of electrical task or mechanical tasks. The predictions obtained from these SMEs showed little differences between courses, unlike the data collected from the course instructors. The causal factors used by the SMEs were different from those used by course instructors, since they were comparing different training devices while the course instructors were comparing training devices to actual equipment trainers.

The use of the CBP procedures made cross-verification a straightforward task. Follow-up telephone conversations focused on factors such as availability and training device learning time. We could ask how long it would take to train students to use the training device, and obtain agreement more easily than if we were questioning the overall judgement. In some cases the follow-up calls had to be made two months after the initial interview. CBP has the virtue of allowing such follow-up communication to occur productively and efficiently.

Furthermore, this has promising implications for the updating of Comparison-Based Predictions throughout the training device procurement cycle. At periodic intervals, as training devices specification proceeds, it should be possible to examine the previous CBP outputs, using the audit trail as illustrated in the results reported for SMEs, in order to modify and improve the predictions.

There are a number of ways that the quality of the CBP data is limited by the nature of the available information. These limitations will affect virtually any prediction effort. Nevertheless it may be useful to describe their impact on CBP.

One major limitation is that there are limited baseline data on the effectiveness of existing training devices, or existing AETs. The lack of an operational data base for maintenance training devices is a severe problem for any predictive strategy. It may preclude the use of formal models. We were able to generate predictions by asking the course instructors to make estimates of the missing training effectiveness data. Basically, this degrades CBP into a structured interview format. However, in the absence of operational data it is not clear that there is any better alternative. Our position is that in the absence of hard data, predictions can only be generated through SME judgement. This can be done informally. CBP is a way to structure the process, and possibly to improve it, although this has not been demonstrated in the current study.

Another problem is that the quality of instructional software has a major effect on the value of a training device, and yet the information we had available for each training device centered around hardware. We did not have any description of the instructional software for the de vices. There is a need for some method for describing instructional capability.

Another issue is the selection of SMEs. Most of the SMEs in this study were course instructors. The value of their judgements can be questioned. It is not our intent to defend their sagacity. Our goal was simply to study whether they could use CBP methods and provide the information needed.

Since CBP depends so strongly on expert judgement;, there is a clear reliance on the identification of qualified SMEs. We have not addressed this problem in this study. We have examined CBP as a way of structuring and improving SME judgement, but we do not have criteria for identifying optimal SMEs. One possible approach is to use different types of SMEs and study the convergence of their predictions, which is an approach we were able to implement.

It is worth noting that along with the general AMTESS recommendations, the CBP approach did generate a specific design recommendation for a different type of training device, one focused on the STE/ICE equipment. Such a TD could be developed using much simpler technology, such as Computer Assisted Instruction (CAI) and graphics, and might be less vulnerable to MOS reorganizations. In addition, cbp also identified some of the weaknesses of this concept (primarily the cost justification), thereby providing evaluation dimensions for making judgements about the value of the recommendation.

In summary, the Comparison-Based Prediction Method was used to collect prediction and recommendation data. The analysis of specific causal factors provided information for data interpretation and verification of estimates. As set of AMTESS recommendations was derived, along with a comparison-based concept for a STE/ICE training device. Lessons were learned about the application of CBP, and suggestions for future uses were made.

It is anticipated that the users for CBP would be personnel involved in training device design and recommendation, who are called upon to generate predictions of the effectiveness of training device concepts. Currently there are no tools for these types of front end analysis predictions. The reliance on subjective judgement occurs by default. If CBP can be used as a means of structuring SME judgement, and thereby improving it, then it can be valuable as a front end analysis tool. The value of CBP can best be established through evaluation research contrasting the predictions made by unstructured SME judgements vs. CBP structured judgements.

REFERENCES

Cicchinelli, L. F., Harmon, K. R., Keller, R. A., Kottenstette, J. P. Relative cost and training effectiveness of the 6883 three-dimensional simulator and actual equipment. Air Force Human Resources Laboratory, AFHRL-TR-80-24, September, 1980.

Cicchinelli, L. F., Harmon, K. R., & Keller, R. A. Relative cost and training effectiveness of the 6883 F-111 converter/flight control system simulators as compared to actual equipment. Air Force Human Resources Laboratory, AFHL-TR-82-30, December, 1982.

Hesse, M. Models and analogies in science. Notre Dame: University of Notre Dame Press, 1966.

Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.

Kane, J. J., & Holman, G. Training device development in the context of Army system acquisition (draft). Behavioral Sciences Research Center SAI Report No. 82-02-178, 1982.

Klein, G. A. Use of comparison cases to predict training effectiveness of maintenance simulators. Klein Associates Technical Report, August, 1982.

Klein, G. A., & Gordon, S. E. Using comparison-based methods for predicting and designing. Proceedings, Psychology in the DoD Symposium, 1984.

Klein, G. A., & Weitzenfeld, J. The use of analogues in comparability analysis. Applied Ergonomics, 1982, 13, 99-104.

McManus, James C. Equipment comparability techniques used during early system design. AFHRL-TR-79-24, JULY, 1979.

NARVA, M. A. Development of a systematic methodology for the application of judgemental data to the assessment of training device concepts. New York, NY: U.S. Army Science Conference, June 1978.

Sternberg, R. J. Component processes in analogical reasoning. Psychological Review, 1977, 84- 353-378.

Swezey, R. W., & Evans, R. A. Guidebook for users of TRAINVICE III. Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences, 1980.

Tetmeyer, D. C. Comparable item approach to esstablishing frequency of maintenance and maintenance tasks for a new aircraft. Wright-Patterson AFB, Ohio: Aeronautical Systems Division, Crew Station, Escape and Human Factors Branch, May, 1976.

Tufano, D. R. The prediction of training device effectiveness: a review of Army models, ARI-TR 613, April, 1982.

Tversky, A. Features of similarity. Psychological Review, 1977, 84, 327-352

Weddle, P. D. Hardman: A total system gestalt. Military Science & Technology, 1981, 1, 44-51.

Weitzenfeld, J. Valid reasoning by analogy: technological reasoning Philosophy of Science, 1984, 51, 137-149.

Widenhouse, W. C., & Romans, W. E. A forecasting technique for operational reliability (MTBF) and maintenance (MMH/FH). Aeronautical Systems Division, Air Force Systems Command, Wright-Patterson AFB, Ohio. Technical Report ASD-TR-77-28, May, 1977.

APPENDIX A

CBP Form, AMTESS

Name______________________________

Office Symbol and Telephone__________________________________

Date_________________ Course Taught_____________________

Students______________ Course Length_____________ A(t)_______

AMTESS device___________________ Task________________________

Comparison device(s)__________________________ B(t)__________

Target Configuration___________________________________

Relative Impact: Target Configuration vs. Existing Configuration:

_____________Better ______________Same __________Worse

Causal Checklist:

	Current Configuration	Target Configuration	AMTESS Hours
Training Potential (Procedures, Perceptual-Motor, Decision making, Task Integration Other			
Utilization (Integration into POI, Ease of Operation, Set-Up time, Performance Evaluation, Instructor Aids, Ease of Modification			
Availability (Reliability, Supportability, Repairability)			
Motivation			
Other			
Total			