Technical Report 652

Human Factors Analysis of Two Prototype Army Maintenance Training and Evaluation Simulation System (AMTESS) Devices

Eleanor L. Criswell, Robert W. Swezey
Science Applications, Inc.

John A. Allen
George Mason University

and

Robert T. Hays
Army Research Institute

Training and Simulation Technical Area
Training Research Laboratory

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

October 1984

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

L. NEALE COSBY
Colonel, IN
Commander

Research accomplished under contract
for the Department of the Army

Science Applications, Inc.

Technical review by

Michael J. Singer
Ray S. Perez

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-TST, 5001 Eisenhower Avenue, Alexandria, Virginia 22333.

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

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

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ARI Technical Report 652

2. GOVT ACCESSION NO.: AD-A158 158

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): HUMAN FACTORS ANALYSIS OF TWO PROTOTYPE ARMY MAINTENANCE TRAINING AND EVALUATION SIMULATION SYSTEM (AMTESS) DEVICES

5. TYPE OF REPORT & PERIOD COVERED: Final Report, June 1983 - October 1984

6. PERFORMING ORG. REPORT NUMBER: SAI 84-11-178

7. AUTHOR(s): Eleanor L. Criswell, Robert W. Swezey (SAI), John A. Allen (GMU), & Robert T. Hays (ARI)

8. CONTRACT OR GRANT NUMBER(s): MDA 903-83-C-0177

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Behavioral Sciences Research Center, Science Applications, Inc., McLean, VA 22102

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2Q263744A795

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: October 1984

13. NUMBER OF PAGES: 66

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):

-- Keywords from p 52

18. SUPPLEMENTARY NOTES:

The appendixes to this report have been published as an Army Research Institute Research Note 84-126.

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

Human factors analysis; Maintenance training devices; Computerized training devices; Instructional features assessment; Generic maintenance training devices; Courseware assessment.

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

A human factors analysis of two prototype Army Maintenance Training and Evaluation Simulation System (AMTESS) devices was conducted. A variety of analytic procedures were applied to assess the degree to which the devices met standards in the areas of hardware, software, and instructional features of the courseware. Both devices met hardware and software standards critical to the performance of their training function, but the courseware did not appear to have been designed around important learning principles. Appendixes to this report have been published as Army Research Institute Research Note 84-126.

to p 52

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

Technical Report 652

Human Factors Analysis of Two Prototype Army Maintenance Training and Evaluation Simulation System (AMTESS) Devices

Eleanor L. Criswell, Robert W. Swezey
Science Applications, Inc.

John A. Allen
George Mason University

and

Robert T. Hays
Army Research Institute

Angelo Mirabella, Contracting Officer's Representative

Submitted by
Robert J. Seidel, Chief
Training and Simulation Technical Area

Approved as technically adequate
and submitted for publication by
Harold F. O'Neil, Jr., Director
Training Research Laboratory

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

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

October 1984

Army Project Number
2Q263744A795

Training 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

This report is part of a larger program to develop new methods for designing and procuring maintenance training devices. In particular, it is one of a series of efforts to help PM-TRADE produce a model for procuring generic maintenance trainers. The report can be used by PM-TRADE as a resource document to evolve general specifications for the Army Maintenance Training and Evaluation Simulation System (AMTESS). It would also be useful in tailoring AMTESS to meet specific future training device requirements.

Finally, the document, though dealing specifically with AMTESS, should also provide a model for conducting human factors analysis and design recommendations for other training devices.

EDGAR M. JOHNSON
Technical Director

ACKNOWLEDGEMENTS

The authors acknowledge the contribution of Martha Cole of Science Applications, Inc., who provided valuable assistance with data reduction.

HUMAN FACTORS ANALYSIS OF TWO PROTOTYPE ARMY MAINTENANCE TRAINING AND EVALUATION SIMULATION SYSTEM (AMTESS) DEVICES

EXECUTIVE SUMMARY

Requirement:

The Army Maintenance Training and Evaluation Simulation System (AMTESS) devices are maintenance training devices which consist of a core component (2-D module) for presenting instruction and a computerized modular 3-D component (actual equipment mock-up) on which the student practices maintenance skills. One AMTESS device was constructed by the Grumman Aerospace Corporation and the other by a consortium of the Seville Research Corporation and Burtek, Inc. The purposes of the present human factors analysis were to expand the AMTESS data base with detailed human factors information, to provide an example of a human factors analysis of a training device, and to offer human factors suggestions for future device design.

Procedures:

Several human factors assessment procedures were devised and/or adapted for use in this assessment and applied to both AMTESS devices. These procedures included a MIL-STD-1472C-based checklist pertaining to hardware and software requirements for military devices, an Analytic Profile System display analysis, an instructional features/courseware assessment, a readability analysis, and a student performance record analysis. (The 200-page checklist, complete with data for both devices, has been published as an Army Research Institute Research Note.) Relevant information previously collected in other AMTESS studies was assembled, summarized from a human factors perspective, and tabulated. Data generated included quantitative measures such as number of critical items from the 1472C-based checklist addressed by each device, and qualitative data such as opinions about useful device features. All information was organized within three broad areas of concern: hardware, software, and instructional features of the courseware. Recommendations were offered concerning future training device design and human factors analyses.

Findings:

In the area of hardware, both devices were found to include all but one feature critical to device functioning as revealed by the MIL-STD-1472C-based checklist. Neither device provides the critical feature of printout of student performance data in the event of power loss. Opinion survey data pointed to poor repair records for both devices.

With respect to software, both devices included desirable software features such as lesson editing capability and malfunction insertion.

Software criticisms include lessons which are difficult to edit and slow system response time for the Grumman device, and the proprietary language used in the Seville/Burtek device.

The results of the courseware evaluation suggested that lesson material and lesson design reflect learning principles only minimally. Both devices present material, accept and evaluate student input, and provide contingent feedback. Neither device allows for repeated practice, presentation of tasks in order of difficulty, varied and multiple presentations of the same material, or systematic review of previously learned skills.

It is suggested that the courseware of future computerized maintenance training devices be based on important learning principles. Allocation of function between device and instructor should be reviewed, and displays should be more carefully composed. Student performance data should be collected and carefully displayed for ease of use.

The results further suggest that a human factors analysis of a training device include components similar to the hardware, software, and courseware assessments discussed in this report. In addition, the courseware and display assessments used in the present analysis might be expanded for use in future analyses. New assessments in the areas of editing capability and device-student interaction are also suggested.

Use of Findings:

These findings will be useful to those interested in human factors analysis of training devices and in training device design. They will help in the development of specifications for the design of future maintenance training devices.

HUMAN FACTORS ANALYSIS OF TWO PROTOTYPE ARMY MAINTENANCE TRAINING AND EVALUATION SIMULATION SYSTEM (AMTESS) DEVICES

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

HUMAN FACTORS ANALYSIS OF TWO PROTOTYPE ARMY MAINTENANCE TRAINING AND EVALUATION SIMULATION SYSTEM (AMTESS) DEVICES

INTRODUCTION

Purpose

This report documents a human factors analysis of two prototype Army Maintenance Training and Evaluation Simulation System (AMTESS) devices. The analysis was conducted to determine the extent to which the AMTESS devices incorporate accepted human factors guidelines (in such areas as hardware design, software capability, and safety) and principles of instructional design (e.g., clear presentation of material, delivery of reinforcement). The conduct and results of this analysis serve three purposes. First, they increase the data base of the AMTESS devices. Second, they serve as an example for conducting a human factors analysis of interactive maintenance training devices. Third, they provide information that will aid in the development of guidelines for the design of interactive maintenance training devices, and of future training device human factors analyses.

AMTESS Devices

The AMTESS devices may be categorized as generic or general purpose training devices. The devices include a two-dimensional (2-D) display which presents textual (words, pictures) training material, and a three-dimensional (3-D) equipment mock-up for practice and testing. This 2-D/3-D combination represents a sophisticated configuration in maintenance training device design (Brock, 1978).

The devices consist of an instructor station (IS) with central computer, a separate student station (SS), and modular 3-D interchangeable components. The devices are general purpose in the sense that the IS and SS have generalized, but incomplete capability, and a modular component tailors the device to a specific application.[1] Thus, one device may be used for as many different training programs as there are 3-D components.

All the military services use general purpose devices. However, AMTESS devices differ from other general purpose trainers in at least one important way: their sophisticated use of high technology with the modular component. When a student manipulates a part of an AMTESS 3-D module (e.g., loosens a bolt, adjusts a dial, removes or installs a part), the action is registered by the computer and its accuracy assessed. Most other devices do not have this capability. In fact, most other general purpose trainers use a 2-D, not a 3-D, component which tailors the device to a specific application.

[1] Orlansky and String (1981) call this type of device "standard."

The two prototype AMTESS devices were constructed to test the technology and were used in small-scale device effectiveness studies. One device was constructed by the Grumman Aerospace Corporation and the other by the Seville Research Corporation and Burtek, Inc.

Figure 1 presents the components of the Grumman device. The IS and SS have separate components but are located at the same desk. The IS includes device control functions with the central computer, videodisc player, dual floppy disk unit, and instructional control functions with a CRT and keyboard (for lesson editing), and a hardcopy printer. The SS contains one component: a CRT where videodisc instructional frames are presented which the student touches to input responses. The 3-D modules (representing the starting and charging system of a self-propelled howitzer engine and the transmitter portion of a high power radar unit), each containing a second computer, plug into a wall socket and attach to the central computer with an RS232 cable.

Figure 2 presents the components of the Seville/Burtek device. The IS and SS are physically separated, each on its own desk. The IS is contained in a desk with three drawers. Device control functions are accomplished with the computer, hard and floppy disk drives, and power supplies, each located in a drawer. The instructional control functions are accomplished with the pushbutton panel (used by the instructor to specify task sequencing within a lesson), CRT with keyboard (for editing any material presented to the student) and hardcopy printer located on top of the desk. The SS contains a pushbutton panel for inputting responses, and two screens all located in one unit. The CRT screen instructs and queries the student. The slide projection screen presents pictures and schematics as needed during training for emphasis. The 3-D modules (a full-scale mock-up of a 5-ton diesel truck engine and the transmitter portion of a high power radar unit) operate off the IS computer, and attach to the IS with four large cables.

Brief History of AMTESS

Criswell, Unger, Swezey, and Hays (1983) present a detailed history of the entire AMTESS project from its inception through device evaluation at two Army schools. Figure 3 presents the timeline of events adapted from Criswell et al. (1983). Briefly, four contract firms (Hughes Aircraft Company, Honeywell System and Research Center, Seville Research Corporation, and Grumman Aerospace Corporation) were commissioned in September 1979 by PM TRADE to conduct front-end anayses (FEAs) and design generic maintenance training devices. Each company submitted designs by July 1980. Still under the auspices of PM TRADE, two firms, Grumman and Seville teamed with Burtek, Inc., each constructed two breadboard prototype devices which were delivered to two Army test schools in 1982.

Evaluation of the prototypes (which began in early 1982) has been a large-scale effort. SAI conducted the transfer-of-training evaluation of the two AMTESS devices. That evaluation found few practical differences between simulator and conventional (actual equipment) training (Unger, Swezey, Hays, and Mirabella, 1984b), and benefits for simulator training in such areas as safety and range of malfunctions used during training (Unger et al., 1984a, c). In addition, school personnel held a positive attitude toward the simulators (Unger et al., 1984a, c). The devices were not

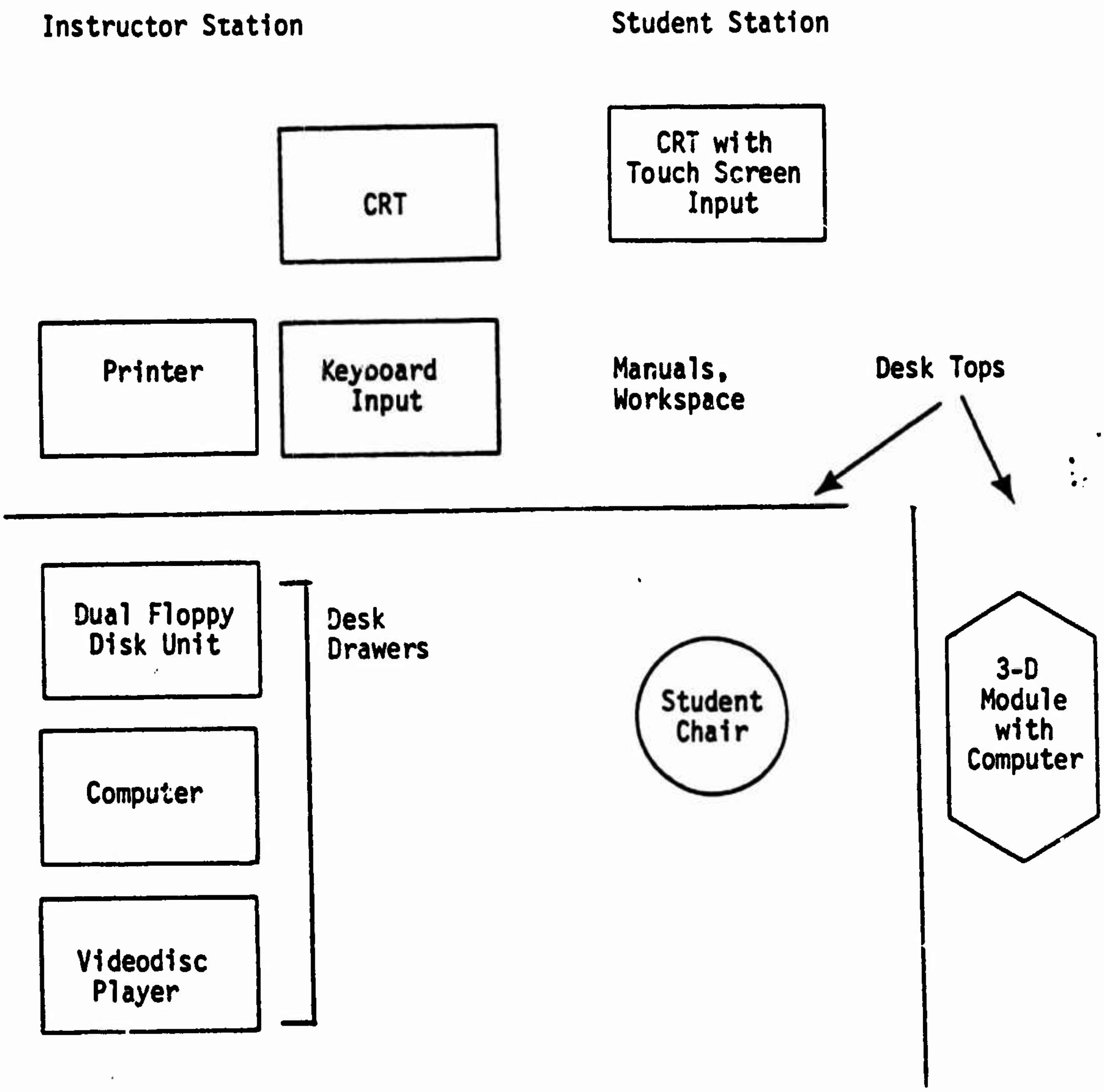

Figure 1. Components of the Grumman AMTESS device.

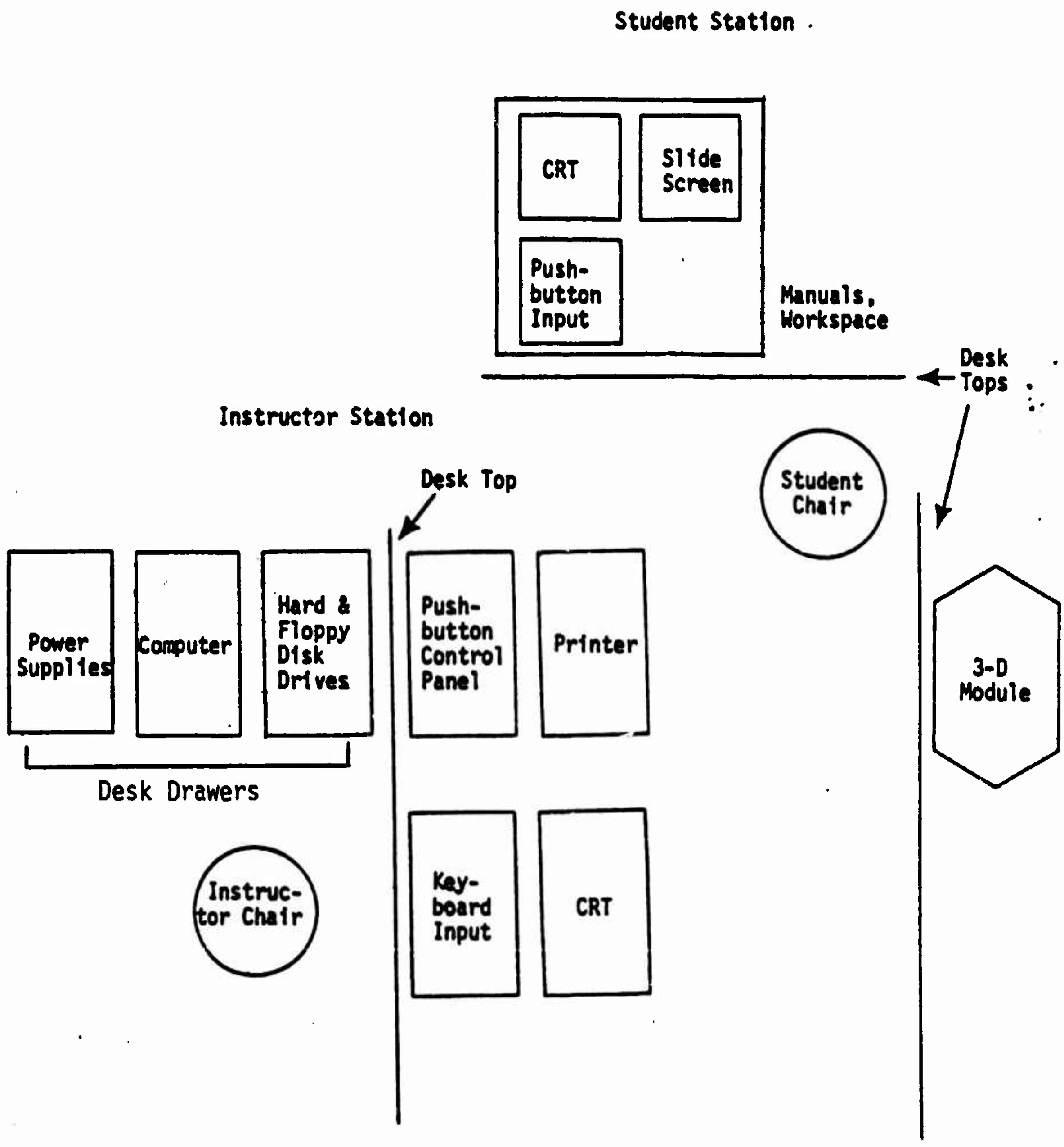

Figure 2. Components of the Seville/Burtek AMTESS Device

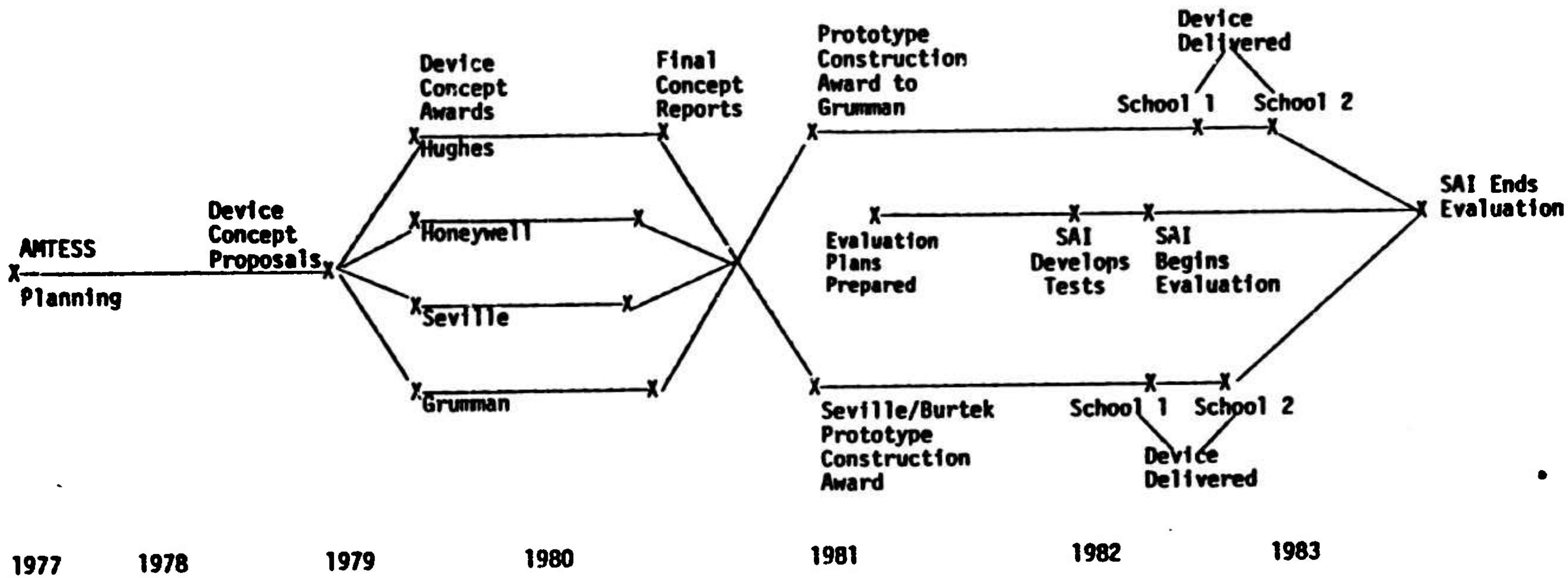

Figure 3. AMTESS milestones (adapted from Criswell et al., 1983, p. 8).

designed or procured for wide-scale use in training, but similar school-ready devices may be procured in the future.

The AMTESS device development process has also been evaluated. Woelfel, Duffy, Unger, Swezey, Hays, & Mirabella (1984) critiqued the FEA procedures used by the four contractors to develop AMTESS concepts. Woelfel et al. (1984) found similarity across all four concepts, adequate FEA documentation for three of the four contractors, but only informal evaluations and acceptance criteria advanced by the government. Criswell et al. (1983) critiqued the entire device development process and found that unclear (or nonexistent) criteria for device acceptance at all stages of development contributed to acceptance of devices with poor operational performance. In addition, government requirements for the devices changed in major ways during device development, but the devices were not changed accordingly, a problem which went unnoticed until the devices arrived at the schools and the evaluation team commenced the evaluation (Criswell et al., 1983). The devices are now housed in a university psychology laboratory. They will be modified for use in basic research on training device features for at least the next few years. This human factors analysis is the final evaluation of the AMTESS devices as they were delivered in 1982 by Grumman and Seville/Burtek.

Human Factors Analyses during Device Development

The development process of a maintenance training device may be thought of as having three stages: device concept development, prototype device development, and production model development. [These stages parallel Kane and Holman's (1982) stages of advanced development, engineering development, and production.] Figure 4 presents the stages in device development and the points at which human factors are a concern. As shown in Figure 4, human factors should be considered by the device contractor at each stage of development. Probably the most thought to human factors should be given during front-end analysis and concept design. Then, after a device is constructed, changes in human factors design will be made as a result of observing real people operate the real device. From the government's perspective, a human factors test should be included in the acceptance testing for each stage in the development. [The U.S. Army is presently working to systematically integrate human factors analysis into its official system development process (U.S. Army, Material Development and Readiness Command, 1984). A developing system would continue to the next life-cycle stage, contingent in part on a satisfactory human factors analysis.]

The government's goal for device concept development is to evaluate alternative device concepts and select the concepts that will become actual devices (Kane & Holman, 1982). Contractors arrive at device concepts by completing front-end analyses. A thorough FEA should consist of seven components as described by Woelfel et al. (1984) in addition to a life-cycle cost estimate. The seven analyses are: task analysis, training requirements analysis, fidelity analysis, reliability and maintainability analysis, organization analysis, human factors analysis, and target population analysis. During FEA, the contractors gather pertinent data, incorporate findings into their design, and submit the design to the government. It is during this time, then, that human factors are first considered by the contractor. The government then assesses the degree to which designs incorporate FEA results including the extent to which human factors principles are incorporated into the design.

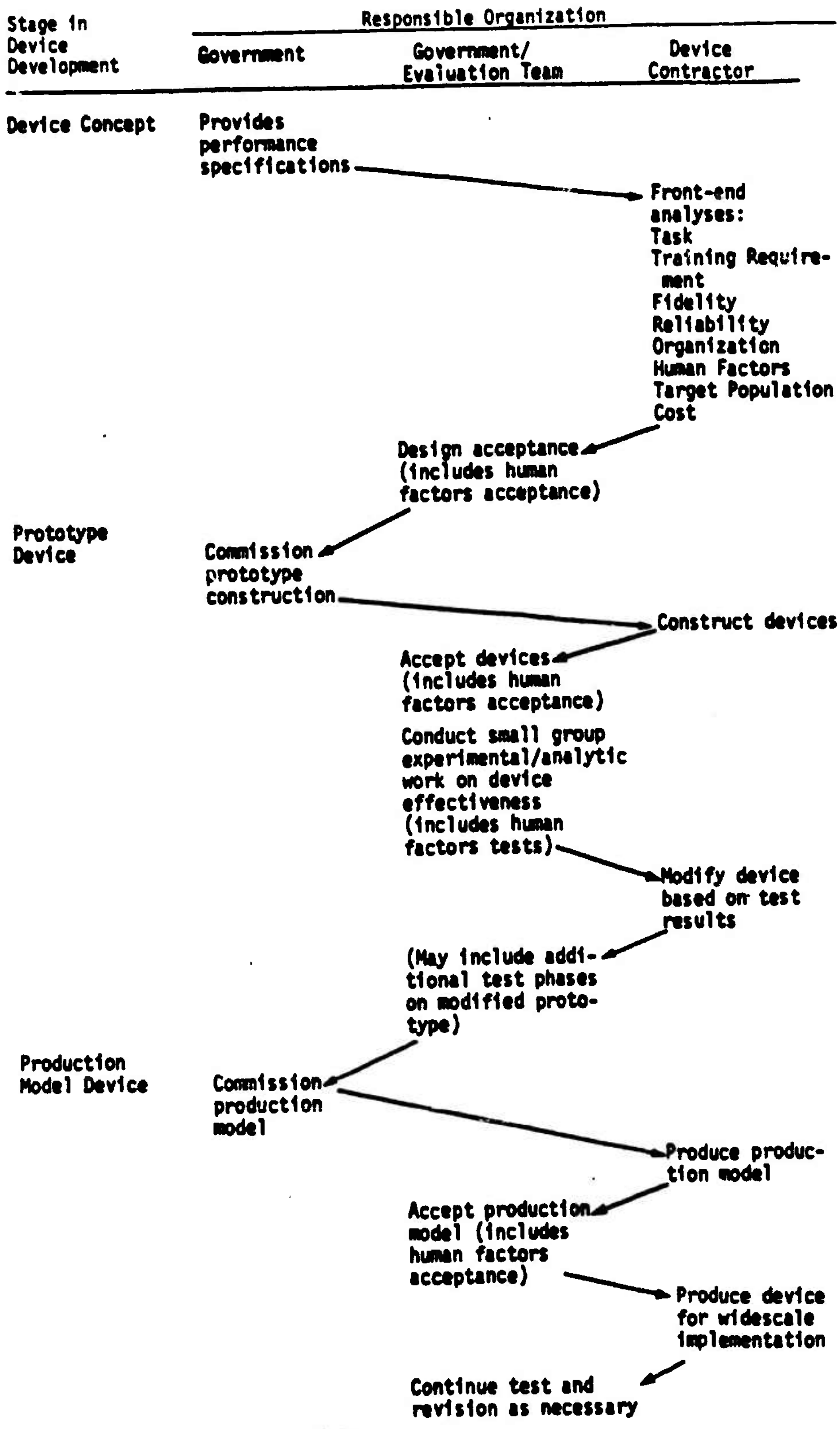

Figure 4. Development process of a maintenance training device.

During prototype development, the contracting firm should not have to repeat its FEA (Kane & Holman, 1982) unless drastic changes in government requirements for the device occur (Criswell et al., 1983). The government, however, should conduct human factors analysis as part of its in-plant review (during device construction) and device acceptance (after construction). These reviews should assess the degree to which human factors and instructional design principles are implemented in the actual device. In addition, the prototype device evaluation team should complete a human factors analysis as part of their evaluation. Usually the object of the first prototype evaluation is to test training effectiveness and transfer of training (Kane & Holman, 1982) and human engineering concerns (Hendricks, Kilduff, Brooks, Marshak, & Doyle, 1983) using a small sample of typical students. Following prototype testing of the first-run or breadboard model, some devices may be modified, a second-run prototype or brassboard model constructed and that model tested (Kane & Holman, 1982). This brassboard step may be omitted if the breadboard functions well or if resource limitations prohibit it.

During the production stage, the contractor and government may conduct analyses to determine if the production model meets specifications and if large-scale use suggests design modifications (Kane & Holman, 1982). This evaluation should include human factors assessment.

During the AMTESS concept development stage, the four contractors completed FEA which included task analysis, training requirements analysis, and fidelity analysis. The influence of these analyses on device design are described in their final reports. The contractors each produced a volume containing the results of the FEAs and a volume presenting "preliminary system engineering design" which consists of schematics and specifications of their proposed devices. Human factors concerns certainly influenced these designs, but no special human factors analysis was required or conducted by the contractors, and the government did not systematically apply human factors criteria in assessing the device designs.

General Approach to the Analysis

There is no one standard procedure for conducting a human factors analysis of a maintenance training device. Smode (1972) presents a broad definition of a human factors analysis, suggesting that a human factors analysis should test the total capabilities of the device as a training tool. Smode suggests this be accomplished by employing two broad types of tests: (1) verification that the device creates a physical and psychological environment adequate for the training required, and (2) verification that use of the device produces learning.

Smode (1972) says that the first type of analysis, physical and psychological environment assessment, is part of the in-plant acceptance process. This analysis appears to have three components. One, it verifies that the approved concept design, based on front-end analysis, is correctly fabricated. Two, it verifies that the fabricated device appears physically compatible with students and instructors. Three, this type of analysis assesses that learning principles are a part of the device, and that the manifestation of these principles is not hindered by physical aspects of the device.

The second type of human factors testing suggested by Smode (1972) takes place on-site and verifies that the device is valid for achieving instructional objectives. The transfer-of-training studies (Unger et al., 1984a, b, c) conducted on the AMTESS devices at the two test sites fall into this second broad type of assessment.

The present analysis pertains to the topics addressed in Smode's (1972) first category of human factors analysis, described above. However, AMTESS devices were not subjected to specific human factors assessments as part of in-plant acceptance because such an assessment is not a formal, separate component of Army acceptance procedures. (It may be in the future, as mentioned earlier.) Thus, the present analysis is the first systematic human factors assessment of the AMTESS devices.

As mentioned previously, there is no standard human factors assessment for a maintenance training device. Therefore, this human factors analysis consisted of a battery of assessment techniques specially chosen for their applicability to the AMTESS devices, considering their stage of device development (i.e., post prototype effectiveness studies). The three broad areas of interest in this human factors analysis are hardware, software, and instructional features of the courseware. This analysis sought to answer the following questions:

1. Do the devices allow for clear presentation of instructional material?

2. How well do the devices and students respond to each other?

3. Is device courseware, or instruction, based on learning principles?

4. Do the devices allow the instructor to change the program?

5. Are displays, labels, signals, and controls easy to use or understand?

6. Are personal safety hazards associated with the devices?

7. Are the devices usually in good working order and easy to repair?

METHOD

Selection of Assessment Instruments

A large number of assessment techniques and instruments were considered for inclusion in the battery used in this human factors analysis. Information about various techniques was found in (1) reports and articles on training devices [e.g., Bailey & Hughes (1980); Caro (1977); Fink & Shriver (1978); Hritz, Harris, Smith, & Purifoy (1980); Klein, Gordon, & Palmisano (in preparation); Orlansky & String (1981)] contained in our literature files, (2) reference books, chapters, and reports related to human factors [e.g., Applied Psychological Services (1970); Chapanis (1976); Davis &

Swezey (1981, 1983); Hendricks et al. (1983); MIL-STD-1472C (1981); Ransey & Atwood (1979); Smode (1972); Systems Exploration, Inc. (1983); Van Cott & Kinkade (1972)]. Only a small number of sources contained checklists or other specific assessments [e.g., Applied Psychological Services (1970); Smode (1972)]. Most sources only contained information which suggested content areas for assessments. From all literature surveyed, a battery was put together which appeared to be applicable for the AMTESS device analysis, considering the stage of device development. In addition, the battery was composed such that information would be obtained about the three broad areas of interest: hardware, software, and instructional features of the courseware. Each instrument selected is described in this section.

It is important to note that of the three components of the devices, IS, SS, and 3-D module, the IS and SS were the focus of this analysis. Given that the 3-D modules are high fidelity replicas of actual equipment, conducting a human factors analysis on the modules would be tantamount to conducting an analysis of actual equipment, a task clearly beyond the scope of this analysis. In addition, much of this information was gathered in the conduct of device FEAs (e.g., Garlichs, Miller, & Davis, 1983). However, some information about the 3-D modules, particularly concerning their maintainability, is presented in this report because of the importance of the information.

Some data relevant to this human factors analysis had previously been collected during other AMTESS studies in the context of several different opinion surveys of AMTESS device experts [Criswell et al. (1983); Klein, et al. (in preparation); Unger et al. (1984a, c)]. The different procedures used by these investigators were seen as offering pertinent and complementary information. The qualitative data in Criswell et al. (1983) were generated by asking respondents to compare, within device, features to each other along four dimensions of value. The Unger et al. (1984c) data were generated by asking respondents to comment on device features, without comparing feature to feature or device to device. The Klein et al. (in preparation) data were generated by asking respondents to compare device to device, feature by feature. All these data were assembled and incorporated into this analysis. The availability of these data rendered unnecessary additional structured interviews and surveys during this analysis. These surveys are described in this section.

Instruments Selected

MIL-STD-1472C-based Checklist. MIL-STD-1472C (1981) is the Department of Defense's 300-page listing of human engineering design criteria for military equipment. All military equipment should be designed in accordance with MIL-STD-1472C. MIL-STD-1472C contains 15 sections of criteria, each section has several subsections, and hundreds of criteria are listed. As MIL-STD-1472C was designed for all military equipment, not just training devices, it was clear that not all of it would be relevant to this AMTESS evaluation. However, many of its criteria are applicable to a maintenance training device.

The process of constructing the 1472C-based checklist consisted of four steps. First, the 15 sections in 1472C were judged for their applicability to the AMTESS devices, and nine sections were found to contain applicable criteria. Controls/display integration (1472C paragraph 5.1) pertains to the coordination between student/instructor actions (movements of device controls) and what is displayed on the IS and SS. Subsections concern general criteria, positional, and movement relationships criteria. Visual displays (1472C paragraph 5.2) pertains to the visual presentation of course-related material to the instructor and student at their stations. Subsections concern general criteria, indicator lamps, CRT, and other displays criteria. Audio displays (1472C paragraph 5.3) pertains to the audio signals presented to the instructor and student at their stations. Subsections concern general criteria, audio warnings, and characteristics of audio warning signals criteria. Controls (1472C paragraph 5.4) pertains to input mechanisms (e.g., keyboard, touch screen) on the IS and SS. Subsections concern general criteria, key operator switches, and linear controls criteria. Labeling (1472C paragraph 5.5) pertains to labels on the IS, SS, and 3-D module. Subsections concern general criteria, orientation and location, contents, qualities, capital vs lower case, and equipment lableing criteria. Workspace design (1472C paragraph 5.7) pertains to physical arrangement of IS and SS. Subsections concern general criteria, standing, and seated operations criteria. Design of maintainability (1472C paragraph 5.9) pertains to IS and SS maintenance, prevention of disrepair. Subsections concern general criteria, mounting of items within units, adjustment control, accessibility, access openings and covers, conductors, connectors, and failure indications and fuse requirement criteria. Hazards and safety (1472C paragraph 5.13) pertains to safety features of the IS and SS. Subsections concern general criteria, safety labels and placards, equipment-related hazards, and electrical/mechanical hazards criteria. User-computer interface (1472C paragraph 5.15) pertains to the ways in which students and instructors contact the device (e.g., CRT, menus) Subsections concern general criteria, data entry, data display, interactive control, feedback, prompts, error management/data protection, system response time, and other requirements criteria.

Six sections of 1472C were judged not applicable to this analysis. The section on anthropometry includes details about hardware design related to body measurements and clothing. It was judged inapplicable because of its detail and because general issues were included in the workspace design section. The environment section pertains to environments (e.g., training rooms) where details about lighting and temperature are important. These concerns are external to the device. The section on design for remote handling pertains to remote controls, not a component of the AMTESS devices. The small systems section pertains to portable equipment such as binoculars, and is not applicable to the AMTESS devices. The sections on ground, shipboard and aerospace vehicles are obviously inapplicable.

Second, potentially applicable criteria within the selected nine sections were restated as questions. Each question was written so that "yes" is the desirable answer. For example, the following is criterion 5.1.1.1 from 1472C concerning controls/display integration:

> The relationships of a control to its associated display and the display to the control shall be immediately apparent and unambiguous to the operator. Controls

should be located adjacent to (normally under or to the right of) their associated displays and positioned so that neither the control nor the hand normally used for setting the control will obscure the display.

This criterion became two questions on the checklist:

- Are relationships of controls to their associated displays, and displays to controls, immediately apparent to the operator?

- Are controls located adjacent to (normally under or to the right of) their associated displays and positioned so that neither the control nor the hand normally used for setting the control will obscure the display?

Thus, there is not a 1:1 relationship between 1472C criteria and items on the checklist. The items, however, constitute a representative sample of concerns based on 1472C which were judged applicable to the AMTESS devices.

Third, other human factors checklists (not designed specifically for training device analysis) were reviewed to determine their relevance to the 1472C-based checklist. Other checklists included: the Army Human Engineering Laboratory guidelines for design of management information systems (Hendricks et al., 1983), a contractor-generated comprehensive supplement to the user-computer interface section of 1472C (Systems Exploration, Inc., 1983), and a comprehensive design guide for the instructor station of a training device (Smode, 1972). All three are based, at least in part, on MIL-STD-1472C and thus did not appear to be useful in generating additional information. It was decided that the scope and level of detail of the 1472C-based checklist was acceptable without additions for our purposes.

Fourth, the checklist was formatted for ease of use, modeled after a human factors checklist used to evaluate a simulated nuclear power plant control room (Malone, Kirkpatrick, Mallory, Eike, Johnson, & Walker, 1980). Figure 5 presents a sample page from the checklist. Space at the top of the form was allocated for identifying the relevant 1472C section, and the device and subsystem under study. The items (questions based on 1472C criteria) were listed down the left column. Four columns were allotted for the answers to the question in each item. The judge, choosing from four answers, "Yes," "No," "Not applicable," and "Unknown, no data," placed an "X" in the appropriate column for each item. In the column labeled "Criticality," the judge assessed the criticality or importance of a "Yes" answer for each item. (The criticality scale is described below.) Finally, space was allotted for comments.

The following criticality ratings, adapted from an Air Force criticality scale (Air Force Systems Command, 1969), were used in this assessment:

1 = not critical to the operation of the training device as a whole or to the individual stations; if the checklist question is answered "No," the device will still be able to perform training as it should.

HUMAN FACTORS ANALYSIS
DESIGN CHECKLIST

AREA CONTROL/DISPLAY INTEGRATION SECTION 5.1 (MIL-STD-1472C)

SYSTEM SEVILLE/BURTEK SUBSYSTEM INSTRUCTOR STATION COMPONENT ______

DESIGN CRITERIA	YES	NO	N/A	UNK	CRITI-CALITY	COMMENTS
5.1 CONTROL/DISPLAY INTEGRATION						
5.1.1 GENERAL CRITERIA						
5.1.1.1. RELATIONSHIP						
Are relationships of controls to their associated displays, and displays to controls, immediately apparent to the operator?	X				3	
Are controls located adjacent to (normally under or to the right of) their associated displays and positioned so that neither the control nor the hand normally used for setting the control will obscure the display?	X				1	
5.1.1.2 DESIGN						
Are control-display relationships apparent to the user through proximity, similarity of groupings, coding, framing, labeling, and similar techniques?	X				2	
5.1.1.3 COMPLEXITY AND PRECISION						
Is complexity and precision required of control manipulation and display monitoring consistent with the precision required of the system?	X				2	
Does control/display complexity match the capability of the operator (in terms of discrimination of display detail), or match the operator's manipulative capability under the dynamic conditions and environment in which human performance is expected to occur?	X				2	

Figure 5. Sample page from the MIL-STD-1472C-based checklist.

2 = important to the function of the IS or SS, and may jeopardize the overall device training function if the item is not satisfied; if not satisfied, causes some degradation in training.

3 = critical to success in training; if item is not accounted for, the device performs at an unacceptable level.

4 = absolutely critical to device functioning; if item is not accounted for, the device will not operate or fail to effect training.

A total of 483 checklist items was applied to the Grumman device, with IS and SS located on one desk. The Seville/Burtek device was evaluated on 612 items: three sections of items were applied separately to the physically distinct IS and SS. For controls/display integration, 16 items were applied to the Grumman device, and applied separately to the IS and SS of the Seville/Burtek device. For visual displays, 53 items were applied to the Grumman device, and applied separately to the IS and SS of the Seville/Burtek device. The audio displays section contained 15 items, applied to each device. The controls section contained 60 items applied to the Grumman device, and the items were separately applied to the Seville/Burtek IS and SS. The remaining sections were applied once to each device: labeling, 44 items; ground workspace design, 9 items; design for maintainability, 61 items; hazards and safety, 9 items; and user-computer interface, 216 items.

Using the checklist, each question was answered for each device. All assessment was completed during April 1984. The Grumman device was operational during the assessment and information was available for 446 of the 483 (92%) items applied. Items where no data were available for the Grumman device primarily concerned the design for maintainability section. Application of those items (e.g., have physical measures been taken to preclude improper mounting of components?) would have required review of detailed design specifications (which do not exist) from the device manufacturer. The Seville/Burtek device was not operational during the assessment, but information was still available for 570 of the 612 (93%) items. As with the Grumman device, some items could not be applied in the design for maintainability section. In addition, no data were available for 21 of the 216 (10%) items in the user-computer interface section (e.g., is a dictionary of abbreviations and codes available on-line?); information would have been available had the device been operational. In summary, the fact that the Seville/Burtek device was not operational had only a minor effect on the data for this particular component of the human factors analysis.

Display Analysis. The Analytic Profile System (APS) (Applied Psychological Services, 1970) is a paper-and-pencil instrument which may be used to evaluate seven characteristics or dimensions of any visual display. The APS is designed for use by human factors experts. The APS contains 35 items, each item is a set of four statements. For each item, the expert judge chooses one statement that best describes the display and one statement that least describes the display. [As reported by Applied Psychological Services (1970), interrater reliability of the APS is .98; concurrent validity with expert judgment is .75; predictive validity using performance occasioned by a display as the criterion has been shown to be .87.]

To score the APS, the expert's selection of best and least descriptive statements for each of the 35 items are tallied with respect to their representation along each of the seven dimensions. Each dimension is represented by 40 statements equally divided among the most and least descriptive statements. A most descriptive statement is scored +1, a least descriptive statement is scored -1. Scores for each dimension are then added. Thus, the score range for each dimension is -20 to +20. Generally, a minus score is poor, zero (0) is intermediate, and a positive score is good. More precise score interpretation requires expert judgment.

The APS yields a score for the following dimensions:

- Volume of material displayed, number of elements in the display
- Format selection vis a vis lesson content
- Signal to noise discrimination
- Organization
- Presentation of main vs. ancillary points
- Overall integration and understanding
- Support to decision making

For purposes of this analysis, the APS was applied to the Grumman device videodisc frames presenting two lessons (set up and check out of the VTM, the electronic box used to diagnose problems in the equipment, and troubleshoot the starter circuit), and applied to 12 randomly selected artwork displays (artist's pictures and schematics) presented on slides by the Seville/Burtek device. (Seville/Burtek text frames could not be evaluated because the device was not operational.) A qualified judge completed both APS applications. Test-retest reliability was .85 for the Grumman frames and .97 for the Seville/Burtek slides.

Multiattribute Utilities Data. Criswell et al. (1983) surveyed four device experts for their quantitative ratings of ten features of the Grumman device and 13 features of the Seville/Burtek device. The data were subjected to a multiattribute utilities analysis which yields a total score for each feature reflecting the value of the feature along several dimensions (Swezey, 1979). Details of the analysis are presented by Criswell et al. (1983). In summary, three steps were taken. First, experts were asked to rank four dimensions on which each feature would be rated. The ranks would become a weight (constant) for each dimension. Second, judges rated the device features on a scale of zero (0) (poor) to 100 (best) along each of the four dimensions. Additional anchors were not provided to the judges. Third, feature ratings were weighted using the constants derived from step one. The process yielded a single score for each feature which reflects its composite value.

The four dimensions used in the analysis were concept behind the feature, implementation or expression of concept, ease of operation of the feature, and motivating quality of the feature. Concept and operations were judged to be the most important dimensions, followed by implementation and motivation.

Table 1 presents the features of each device which were rated in the analysis.

Table 1

Features Rated by Judges in Criswell et al. (1983)

Grumman Device	Seville/Burtek Device
3-D module	3-D module
Student performance record available in hardcopy	Student performance record available in hardcopy
Instructor CRT	Instructor CRT
Videodisc	Slide projector unit
Touch panel, student CRT	Student response panel
Editing system for resequencing lesson, available to instructor	Editing system for resequencing lesson, available to instructor
Request help option available to instructor	Student CRT
Repeat lesson option available to student	Instructor control panel
Mandatory instructor call after 2 errors, required of student	Remove/replace capability
Performance feedback, message from computer to student during training	Random malfunction selection
	Performance feedback to student during training
	Troubleshoot only mode
	Sound effects

Qualitative Comments. Qualitative comments about device features were available from two sources, questionnaires and interviews with personnel involved in the transfer-of-training studies (Unger et al., 1984c), and interviews with key personnel in the development of the AMTESS devices (Criswell et al., 1983). Summaries of these previous efforts were compiled for use in the AMTESS human factors evaluation.

Unger et al. (1984c) gathered and summarized qualitative data which relate to a human factors analysis in the context of:

- written instructor questionnaire which included questions that would elicit comments about device features (e.g., "Why do you like or dislike the simulator?"),

- written trainee questionnaire which included questions about specific device features (e.g., "What specific features of the simulator do you like or dislike?"), and

- oral interviews with various knowledgeable people which included questions about specific device features (e.g., "Which features helped make the lessons interesting to students?").

Criswell et al. (1983) report qualitative data about device features gathered from structured oral interviews with contractors, device manufacturers' representatives, school (test-site) personnel, and the government. An example of questions which produced data relevant to the human factors analysis include "Why was videodisc chosen?" and "What features made this simulator difficult to operate?". These data were summarized and added to the human factors analysis data base.

Feature Comparisons. Klein et al. (in preparation) collected subject matter expert opinions about AMTESS device features while implementing a paper-and-pencil training-effectiveness evaluation (called Comparison-based Prediction). Two subject matter experts directly compared the two AMTESS devices, feature by feature, and selected the device with the better feature. The features compared were: 3-D module, student performance record, instructor CRT, student CRT, visual display, student response panel, editing system, student request help, and performance feedback during training. In addition, qualitative comments about device features (hardware, software, instructional support, and availability) were obtained from ten school personnel who compared the AMTESS devices to actual equipment trainers, to each other, and to an Air Force maintenance training device (the 6883). The Klein et al. (in preparation) data were seen as augmenting similar data in Unger et al. (1984c) and Criswell et al. (1983); their data were summarized and added to the present analysis.

Courseware Assessment. For purposes of this human factors analysis, instructional features of the courseware were grouped into three categories:

- Features that enable the device to teach according to principles of learning and behavior ("Learning Principle Features").

- Features that enable the instructor to vary instructional material and consequences (or feedback) presented to the student ("Instructor Features").

- Features which improve the flow of training, but which are not based on learning principles ("Optional, but Helpful Features").

The first type is the most important, because without these features, teaching, and thus learning, will be haphazard. These features should be programmed into the device. The second type is next in importance -- it allows the instructor to intercede in training and make improvements in training over that which the device does automatically. Type two, instructor features, are really useless without learning principle features; although an instructor might be able to override deficiencies in the instructional plan of the device by using emphasis features, it would be time consuming and tedious and would negate the benefits of a computerized training device. Finally, type three or "optional" features are nonessential, but they improve the flow of training and allow the device to perform some functions which the instructor would either perform manually or omit.

Instructional features assessment of the courseware, especially including the first two types of features, is important in a human factors analysis. Issues from the last nine years in the journals, Human Factors and Educational Technology, were reviewed for information on courseware evaluation, and of the few articles found which addressed this topic (e.g., Cohen, 1983), none concerned the assessment of instructional features. This appears, then, to be a novel component in human factors analysis of training devices. A checklist was developed to assess the extent to which AMTESS devices included all three types of instructional features. First, features related to learning principles were described, and checklist items developed relating to each principle. It is outside the scope of this report to discuss in detail the principles of learning and behavior. [The interested reader may consult Catania (1979), Powers and Osborne (1975), and Skinner (1968) as references on learning principles.] However, it is widely accepted that learning proceeds in an orderly fashion, and that effective training devices should be based on learning principles. In the basic learning paradigm, instructional material is presented, the student responds to it, and the device delivers a consequence to the student based on the response. A training device presents material, usually on a CRT, the material requires the student to respond (with a command such as "Enter your choice now," or "Connect the wire as you have learned"), the student enters a response (e.g., via keyboard, touch screen, 3-D module manipulation), the device checks the response for accuracy, and delivers the appropriate consequence (e.g., "That's correct," or "Try again"). Incorporated into this type of instructional delivery system are the kinds of learning principles described below.

Reinforcement means that the consequence delivered increases the probability that the student's response will occur again, i.e., that the correct step will be taken if presented again with the same task. Examples of features related to reinforcement delivered by a training device include favorable comments on the CRT and pleasant tones. In addition, more natural reinforcement is derived from such things as actually finding a fault or successfully removing a part. The effect opposite reinforcement is punishment. Punishment means that the consequence delivered decreases the probability that the student's responses will occur again, i.e., that an incorrect action will not be repeated under the same circumstances. Only very mild punishing consequences should be delivered by the device. On a training device, punishment takes the form of informing the student an error has been made. This may be done by a message (such as "Try again"), an unpleasant tone, or a brief pause ("freeze"). In addition, more natural punishers are delivered by such things as parts that will not fit together if improperly aligned.

Learning principles concerning shaping, chaining, prompting, and fading should be considered in courseware evaluation. Shaping is the process of gradually raising the criterion for reinforcement as the student becomes more proficient. On a training device, shaping is accomplished by gradual increase in task difficulty. Chaining is the process of teaching a student to perform one step, followed by another step, followed by another, and so on until the student performs a chain of uninterrupted responses. If a sequence has five steps, each step is taught one at a time, then practiced two at a time, then three at a time, and so on until the sequence is performed without interruption. In a chain, the completion of one step signals the next step. Reinforcement comes after the last step. Devices should allow for the development of chains, which are the target performances. Prompting and fading refer to the procedure of giving extra help to prime the student to respond to instructional material. As the student learns, prompts are faded or gradually dropped out of the training. Prompting and fading may be programmed into the device, or prompts may be solicited by the student using a "request help" feature.

The principle of stimulus control should be considered in courseware assessment. Stimulus control is the process by which the student learns to perform a particular response when presented with a certain circumstance. A device accomplishes stimulus control by allowing for frequent practice of each skill taught, where reinforcement is delivered following only correct responses. Two related processes are stimulus and response generalization. Stimulus generalization is the process by which a student learns to perform a particular response when presented with any number of similar circumstances. Stimulus generalization is prerequisite to transfer of training to the actual equipment. A device programs generalization by presenting the same problem in different ways (e.g., using several different schematics and pictures to illustrate the same problem) and by teaching principles such as a general approach to troubleshooting or engine design. Response generalization is the process by which a student learns to perform several similar responses when presented with a certain circumstance. The device programs response generalization by teaching general skills such as inspecting or listening (for signs of equipment malfunction).

Next, "instructor" features were listed as items on the checklist. Many instructor features relate to learning principles. Also included is a "select next activity" feature which allows the instructor to select the student's next activity rather than taking the programmed next one in sequence. Other instructor features are a control for rate of presentation of material which allows the instructor to speed or slow pace, and a means for message adjustment which allows the instructor to vary the messages (feedback) presented to students.

Finally, "optional features" were listed as checklist items. They include the following helpful, but not critical features: hardcopy printout of student performance record where data are available on CRT, storage of individual student records, and a means to select variables available on the student performance record.

The courseware of both devices was assessed using the instructional features checklist. For the Grumman device, the checklist was completed by reviewing the first two Grumman lessons, a script of the Grumman videodisc frames, and the Grumman instructor and student handbooks. The Seville/Burtek device was not operational and there is no script of its text, so the checklist was answered based on the lesson information contained in the student and instructor handbooks. Where more information was needed about courseware for either device, a knowledgeable staff member was consulted. Nearly all items were answered for both devices.

Readability Analysis. Each device is accompanied by instructor and student booklets. These booklets include: Seville/Burtek Instructor's Guide which describes the simulator, training management, and operating instructions; Seville/Burtek Student Handbook which describes the simulator, introduces the actual equipment, and outlines each lesson in the course; Grumman Instructor Handbook which describes the instructor procedures and instructor options; and Grumman Student Handbook which describes the device, basic computer procedures, and introduces training. All documents are designed to be read before training begins and used as necessary during training. Documents of this type serve the important function of providing an overall perspective on the training that will follow, and they must be understandable. In fact, in some training programs, mastery of handbook material is a prerequisite for progressing to the lessons. Although no explicit requirements (including no readability requirement) concerning these documents were set forth for AMTESS materials, examination of the documents was considered a relevant aspect of this human factors analysis because this analysis was to serve as a sample human factors analysis.

Several samples from each of the four types of documents were subjected to a Fog readability analysis to determine the reading grade level of the material. To calculate the Fog Index:

- Take a passage of about 100 words.
- Divide the total number of words by the number of sentences in the passage. This is an average sentence length.
- Count the words having three or more syllables. (Do not count proper nouns, hyphenated words, or verbs with suffixes.)
- Add the average sentence length to the number of three syllable words.
- Multiply this sum by .40 (four-tenths). This is the Fog Index, which is a quick, reasonably valid reading grade level for that passage.

As noted by Sticht (1969), the reading grade level of some technical material may be unfairly elevated by the presence of three-syllable technical words which are, in fact, part of the reading vocabulary of the user. Such words (e.g., "carburetor") were noted in the readability analysis.

Student Performance Record Analysis. Student performance data collected by the devices were not used in the AMTESS device evaluations because the instructors involved were not required to keep records of this kind or to use the data to monitor or improve training. Performance data used by the device evaluation team were designed and collected manually by the team. However, the devices were required to demonstrate this capability, and under normal circumstances, student performance data are important to a training course. For example, using data such as number of errors and time spent on a lesson, an instructor (or a computer) could decide if a student had mastered one lesson and was ready for the next. To make decisions about the overall course (e.g., is Lesson 2 associated with frequent errors, is Lesson 3 too easy), the same data could be reviewed for several students.

In general, for day-to-day decisions, minimal requirements include information such as name or ID number, date, time of day; lesson or step name; time elapsed in lesson or step; number of errors in lesson or step; and number correct in lesson or step. Information should be carefully formatted and written in understandable English for ease of use. In addition, summary information (including charted data) about a student's performance across steps or lessons is helpful for making progress decisions about any one student. For overall course decisions, information should be retrievable for groups of students by lesson (or step). Charted data is also helpful in this regard. Because of the importance of student performance data, printouts of student records from both devices were assessed for their understandability and comprehensiveness according to standards described above.

RESULTS

MIL-STD-1472C-based Checklist

The completed checklists comprise nearly 200 pages and are available in two appendixes as an Army Research Institute Research Note. A large number of items on the checklist were scored Not Applicable or Unknown/No data (132 for Grumman, 178 for Seville/Burtek). Reasons for scoring an item as "unknown" were described earlier. Items receiving either of these scores were not included in the analysis; therefore, the analysis is based only on applicable items (the total number of items minus the number of Not Applicable and Unknown items), 351 for Grumman and 434 for Seville/Burtek.

The checklist data may be used to address the extent to which the devices contained features at all four levels of criticality. Tables 2, 3, and 4 present this information for both devices. Table 2 concerns features which are critical to the teaching function of the device, that is, those features which received a criticality score of 4 or 3. As shown in Table 2, only a small number (12/351 for Grumman and 18/434 for Seville/Burtek) of items are considered critical to device functioning. Both devices addressed the few critical (scored 4 on criticality) items and all but one of the items scored 3 -- neither device provides hardcopy printout of student performance data in the event of power loss.

Table 3 presents a list of the features which were determined to be critical to both devices. A feature is considered critical if it was scored 3 or 4 on criticality on the 1472C-based checklist. An additional feature, a printer for student performance data printouts was critical for the Grumman device because that information is not available on a CRT. It is

Table 2

Number Critical Items Satisfied by Both Devices

1472C-based Checklist Section	Grumman Device			
	# Items Scored 4[a]	# Satisfied	# Items Scored 3[b]	# Satisfied
Control/Display Integration:				
Overall	0	0	1	1
Instructor Station	-	-	-	-
Student Station	-	-	-	-
Visual Display:				
Overall	1	1	7	7
Instructor Station	-	-	-	-
Student Station	-	-	-	-
Audio Display	0	0	0	0
Controls:				
Overall	-	-	-	-
Instructor Station	-	-	-	-
Student Station	-	-	-	-
Labeling	0	0	0	0
Workspace Design	0	0	0	0
Design for Maintainability	0	0	0	0
Hazards and Safety	0	0	0	0
User-Computer Interface	1	1	3	2

[a]Most critical items, scored 4 on criticality scale

[b]Second most critical items, scored 3 on criticality scale

CONTINUED ON NEXT PAGE

Table 2

Number Critical Items Satisfied by Both Devices, Continued

1472C-based Checklist Section	Seville/Burtek Device			
	# Items Scored 4[a]	# Satisfied	# Items Scored 3[b]	# Satisfied
Control/Display Integration:				
Overall	-	-	-	-
Instructor Station	0	0	1	1
Student Station	0	0	1	1
Visual Display:				
Overall	-	-	-	-
Instructor Station	1	1	6	6
Student Station	1	1	5	5
Audio Display	0	0	0	0
Controls:				
Overall	-	-	-	-
Instructor Station	0	0	0	0
Student Station	0	0	0	0
Labeling	0	0	0	0
Workspace Design	0	0	0	0
Design for Maintainability	0	0	0	0
Hazards and Safety	0	0	0	0
User-Computer Interface	1	1	3	2

[a]Most critical items, scored 4 on criticality scale
[b]Second most critical items, scored 3 on criticality scale

Table 3

1472C-based Checklist Features Critical to the Teaching Function of Both AMTESS Devices

CONTROL/DISPLAY INTEGRATION
• Control-display relationships are immediately apparent.
VISUAL DISPLAYS
• System mode or condition is clearly displayed.[a]
• Alerting/warning displays improve detection of conditions.
• Only relevant information is displayed.
• Displayed information is as precise as necessary.
• Displayed information is immediately usable.
• Information remains displayed long enough to be detected.
• Mentally translating displayed information is unnecessary.
USER-COMPUTER INTERFACE
• Programs and user input/output features allow users to use the system for its intended purpose.[a]
• Programs respond quickly enough, and provide enough detailed and precise information to accomplish the system's purpose without stressing the user.
• Hardcopy printout of student performance data is available where data could be lost by power interruption (only feature not included in either device).
• Hardcopy printout of student performance data is available where record keeping is required.

[a]The most critical features.

Table 4

Number Less Critical Items Satisfied by Both Devices

1472C-based Checklist Section	Grumman Device			
	# Items Scored 2[a]	# Satisfied	# Items Scored 1[b]	# Satisfied
Control/Display Integration:				
Overall	10	10	5	4
Instructor Station	-	-	-	-
Student Station	-	-	-	-
Visual Display:				
Overall	12	10	27	25
Instructor Station	-	-	-	-
Student Station	-	-	-	-
Audio Display	6	5	5	4
Controls:				
Overall	0	0	48	37
Instructor Station	-	-	-	-
Student Station	-	-	-	-
Labeling	15	15	26	20
Workspace Design	0	0	9	9
Design for Maintainability	5	5	19	15
Hazards and Safety	1	1	5	3
User-Computer Interface	73	58	72	64

[a]Important but not critical items, scored 2 on criticality scale

[b]Not at all critical, may be helpful, scored 1 on criticality scale

CONTINUED ON NEXT PAGE

Table 4

Number Less Critical Items Satisfied by Both Devices, Continued

1472C-based Checklist Section	Seville/Burtek Device			
	# Items Scored 2[a]	# Satisfied	# Items Scored 1[b]	# Satisfied
Control/Display Integration:				
Overall	-		-	-
Instructor Station	10	10	3	3
Student Station	10	10	3	3
Visual Display:				
Overall	-		-	-
Instructor Station	14	14	20	19
Student Station	11	10	13	13
Audio Display	9	9	4	3
Controls:				
Overall	-	-	-	-
Instructor Station	0	0	39	34
Student Station	0	0	28	22
Labeling	16	15	26	22
Workspace Design	0	0	9	9
Design for Maintainability	4	1	32	28
Hazards and Safety	1	1	6	5
User-Computer Interface	83	80	74	68

[a]Important but not critical items, scored 2 on criticality scale

[b]Not at all critical, may be helpful, scored 1 on criticality scale

interesting to note that all critical features fell in only three of the nine checklist sections, control/display integration, visual display, and user-computer interfaces.

Table 4 concerns features which are not critical to device function, that is, those features receiving a criticality score of 2 or 1. Table 4 shows that most of the items (96% for Grumman and 95% for Seville/Burtek) on the 1472C-based checklist are less critical items. The absence of any one of these items does not affect the device, but the absence of combinations of less critical items may degrade device effectiveness. Thus, it is important to consider how well the devices addressed these items.

Table 5 presents the percent of the total number of applicable items addressed by both devices on each checklist section. Table 5 shows that the Grumman device addressed most of the items applied to it. The device satisfied the largest percentage of items in workspace design, visual displays, and control/display integration. It scored a middle range on items related to labeling, design for maintainability, user-computer interface, and audio displays. The device addressed the fewest items on the controls and hazards section. The effects of these omitted items on the functioning of the Grumman device are described below.

For visual display, two main deficiencies with the Grumman device were noted. One, a student CRT lockup due to student errors and/or equipment malfunction is not immediately apparent to the user, so the user continues to touch the screen to input data after touching is no longer functional. Two, the short CRT cables prevent improving the visual angle between screen and student; the screen has a hood which prevents full screen viewing unless the student is seated.

The Grumman audio display (a tone or chirp) was reported to be annoying and too long. A more serious problem, however, is that the same tone is used to signal two different things: (1) look at the CRT, and (2) you have made an error. Two different stimuli are needed.

The items not addressed by the Grumman device in the controls section of the checklist were not important, and mainly concerned the on-off lock apparatus and key.

For labeling, three main problems were noted with the Grumman device. Some labels are obscured, some stenciled labels have quickly worn off, and the location of a label with respect to the part labeled is not always consistent.

The most important deficiency seen with the maintainability design items is vulnerability of some Grumman cable connectors due to their exposed location and sensitive construction.

The Grumman device failed to address an important safety item concerning the squared, sharp corners on drawers on the 2-D module, particularly the shin-high videodisc drawer which is left open during training. Numerous leg injuries (gashes, bruises) have been blamed on those corners.

Table 5

Percent Applicable Items Addressed by Both Devices

1472C Based Checklist Section	Grumman Device Number Applicable Items	Grumman Device Percent Items Addressed	Seville/Burtek Device Number Applicable Items	Seville/Burtek Device Percent Items Addressed
Control/Display Integration				
Overall	16	88	-	-
Instructor Station	-	-	14	93
Student Station	-	-	14	93
Visual Display				
Overall	47	91	-	-
Instructor Station	-	-	41	98
Student Staion	-	-	30	97
Audio Display	11	82	13	92
Controls				
Overall	48	77	-	-
Instructor Station	-	-	39	82
Student Station	-	-	28	75
Labeling	41	83	42	90
Workspace Design	9	100	9	100
Design for Maintain-ability	24	83	36	81
Hazards and Safety	6	67	7	86
User-Computer Inter-face	150	83	162	94

The user-computer interface section contained the largest number of items, and many were not addressed by the Grumman device. While the items are not critical, they suggest areas of improvement needed in the Grumman device. These items include: unclear messages and codes, especially in the authoring system; sluggish feedback to student; no explanation about delays (due to computer processing speed) presented to user; no explanation for rejected input (from student or instructor) presented to user; and mass storage of student performance data needed.

Table 5 also shows that the Seville/Burtek device, like the Grumman device, addressed most of the applicable, largely less critical, items. As shown in Table 5, all items were addressed in the workspace design section; visual display items were nearly all addressed. Several sections of items fell in a middle range: user-computer interface, control/display integration, audio displays, labeling, and hazards. Lowest scores on percent items addressed were obtained for the controls and design for maintainability sections. No major problems occurred due to items not addressed in the control/display integration, visual display, audio display, or hazards and safety sections of the checklist.

For controls, on both instructor and student stations of the Seville/Burtek device, use of controls (especially the required sequence of control use) must be taught. Control use is not obvious or clear from the perspective of the consoles. Control labels are unclear on the student input panel; the meaning of buttons labeled "test," "service," and "inspect" is unclear.

For labeling, some improvements are needed on the Seville/Burtek device, particularly on parts related to device maintenance and repair. Abbreviations are used on labels and may be accidentally removed.

For design for maintainability, some items not addressed by the Seville/Burtek device were important. The device should be packaged in a more modular way to assist repair personnel. Some device components (e.g., the backplane section of the computer) are relatively inaccessible and would be difficult to repair. The large cables connecting the 2-D and 3-D modules are vulnerable to damage (resulting from accidental kicking or stumbling) because they are unprotected and lie on the floor.

For user-computer interface, several important items not addressed by the Seville/Burtek device were noted. These include: need for on-line help to recover from errors; need for simplified utility programs; no procedure available for orderly shutdown in the event of computer failure; clearer prompts needed; and wider range of data manipulation options needed.

<u>Display Analysis</u>

Figure 6 presents the Analytic Profile System (APS) display analysis scores for the Grumman videodisc displays of instructional material from two lessons. Taken as a whole, it appears that the Grumman display scores in the intermediate range of display quality; the display is not particularly good or bad. It received its highest score on format reflecting that its

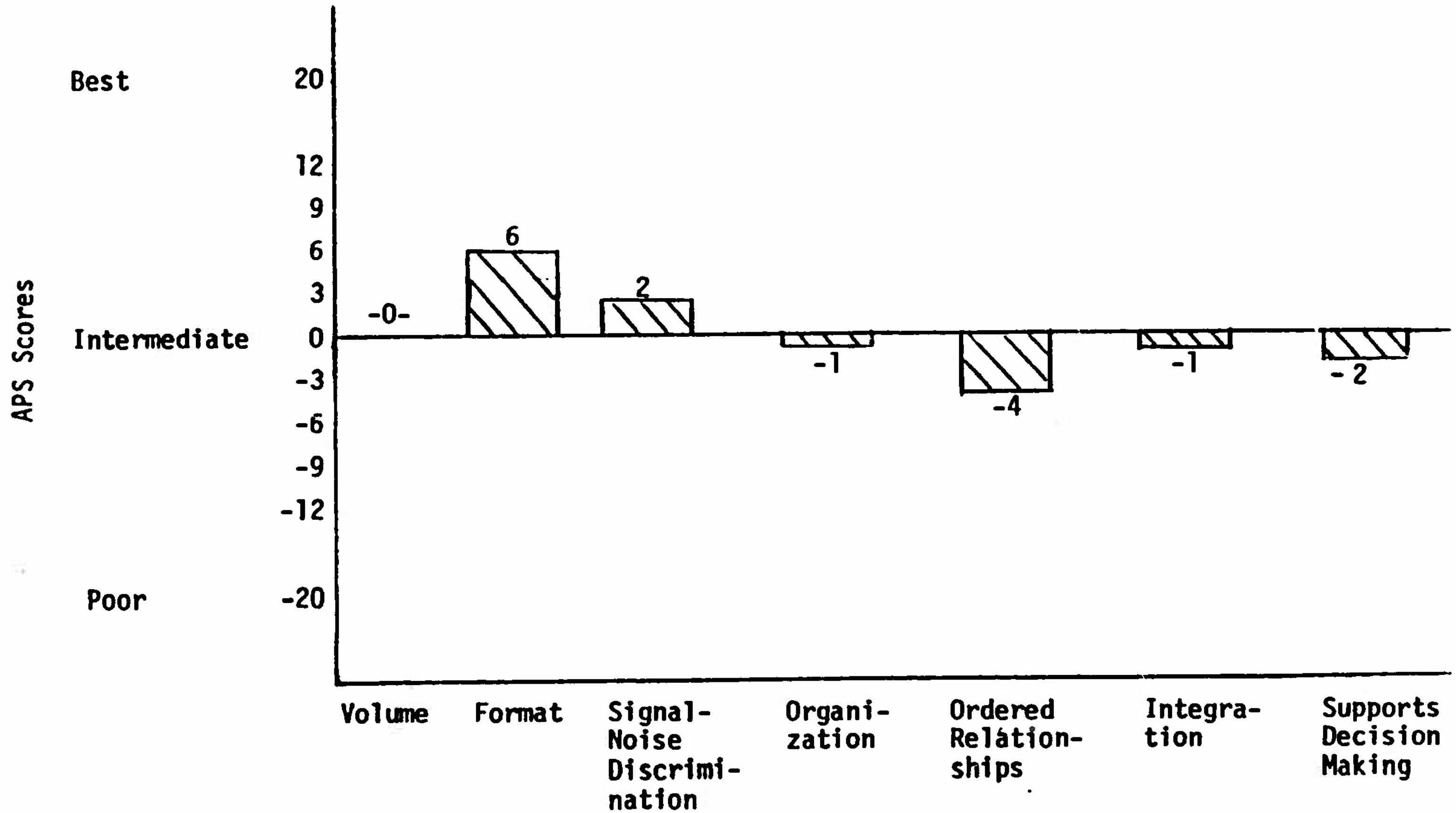

Figure 6. APS scores on seven aspects of the display used in Grumman videodisc instructional frames.

format of presentation (e.g., blue text frames, consistency in query frames, consistency in feedback frames, action sequences for demonstrations) is appropriate, and is the strong point about the Grumman display. As suggested by its signal/noise discrimination score, the display also does an adequate job of clearly presenting the main point.

Three aspects of the display hover around the middle quality level. These three aspects may merit attention in the future. One, volume of information presented per display is about as high as it can get, especially on the text slides. The large volume is supported well, however, by consistency in format. Two and three, the negative scores on organization and integration, probably mean that the volume of information presented per display requires extra attention from the user to organize, read, and integrate the information. Particularly for the text slides, short summaries (and reduced volume per frame) might improve the display.

The Grumman display scored lowest on decision-making support and ordered relationships. The decision-making score suggests that the problems in ordered relationships affects the overall interpretation of the slides. The score on ordered relationships means that some displays caused problems for the viewer in perceiving relationships among displayed items. This score probably reflects the large amounts of words presented, especially on text slides. The text slides might have highlighted key points or included short summaries as mentioned above.

Figure 7 presents the APS scores for the display used in Seville/Burtek slides. The slides contained artwork drawings and schematics of the simulated equipment. They were an adjunct to instruction, displayed only to clarify or emphasize. Thus, the display has a much different purpose than did the Grumman display evaluated. As a whole, the scores in Figure 7 suggest that this display was of intermediate quality, not especially good or bad.

The two strong points of the drawings were organization and format. The drawings were of two types, art and schematics, and the format of all slides within a type remained consistent. The positive signal/noise score probably reflects the artwork's clear labeling with a red arrow pointing to the important parts.

Four scores were negative, suggesting improvements should be considered. The slides might be improved by a reduction in volume of information presented per slide, particularly on the schematics. The negative decision-making score probably reflects difficulties in reading the labels on the schematics; the type-size used was very small. The negative score on integration probably reflects problems in using the detailed, cluttered schematics. Finally, the display scored low on ordered relations indicating that the displays, especially the schematics, did not clearly distinguish principles from secondary relations.

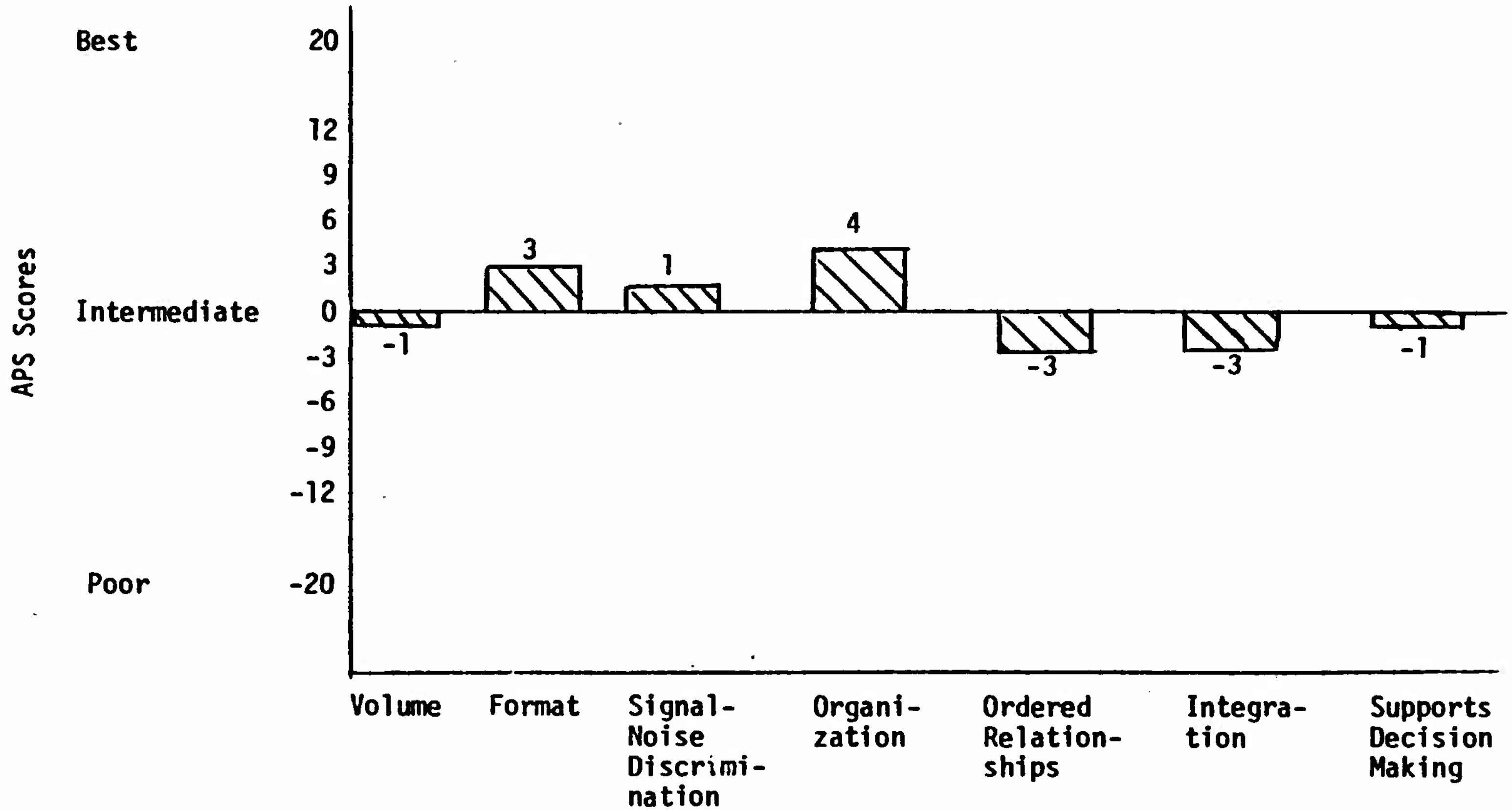

Figure 7. APS scores on seven aspects of the display used in artwork on Seville/Burtek slides.

Multiattribute Utilities Ratings

Table 6 presents the device features rated in Criswell et al. (1983), listed in descending order of composite value. Table 7 presents the rank order of each device feature along each dimension of value. Table 6 presents composite value and Table 7 presents component value. The ranks in Table 6 reflect weighted scores and the ranks in Table 7 reflect unweighted scores.

As described in Criswell et al. (1983), for the Grumman device, the 3-D module received the highest scores for design and performance. Five features fell in a middle range. The repeat lesson option, request help, touch panel, and performance feedback features were viewed as adequate in design and performance. The videodisc was viewed as highly motivating and excellent conceptually, but had some problems practically. Two features seemed in special need of improvement: the mandatory instructor call after two errors and the student performance record. The former was felt to foster dependence on the instructor and perhaps decrease motivation. Further, operational problems were associated with the feature. The student performance record was seen as inadequate because so few student progress data were presented.

For the Seville/Burtek device, the 3-D module and student performance record received the highest scores for design and performance. The device's troubleshoot only mode also received praise as a valuable and dependable device feature. Several features fell into a middle range. Those features are performance feedback, remove/replace capability, student CRT, and random malfunction selection. In the opinion of interviewees and experts, those features were designed well and performed adequately. However, two features were found to be especially in need of improvement. Those features were the student response panel and slide projector unit. In general, the opinion was that the design of the response panel might be improved. The main problems with the slide projector unit were operational ones.

The results from this earlier analysis are augmented by the results from qualitative data collection efforts described below.

Qualitative Comments

The qualitative data in Unger et al. (1984c) were reanalyzed from a human factors perspective. Table 8 presents a summary of qualitative comments related to human factors concerns made by instructors, students, and device experts. For the Grumman device, the videodisc, instructor pre-lesson system check, performance feedback to students, instructor malfunction insertion features, and student request help feature were seen as especially strong points of the device. Weaknesses of the device included: poor repair record, poor student performance record (only available on hardcopy), slow system response time, and the difficult procedures required for an instructor to insert and remove malfunctions. For the Seville/Burtek device, the main assets were seen as its flexibility, fairly good repair record, and the ease with which instructors insert malfunctions. Its two main liabilities were the complicated procedures the student had to follow to interact with the computer, and the low durability of parts on the 3-D module.

Table 6

AMTESS Device Feature Ranks Based on Composite Multiattribute Utilities Scores (adapted from Criswell et al., 1983, p. 36, 46)

Rank	Grumman Device Feature	Composite Feature Value
1	3-D module	Most Valued
2	Performance feedback	↓
3	Request help	
4	Touch panel	
5	Videodisc	
6	Repeat lesson option	
7	Student performance record	
8	Call instructor after two errors	Least Valued

Rank	Seville/Burtek Device Feature	Composite Feature Value
1	Student performance record	Most Valued
2	3-D module	↓
3	Troubleshoot only mode	
4	Performance feedback	
5	Remove/replace capability	
6	Student CRT	
7	Random malfunction selection	
8	Student response panel	
9	Slide projector unit	Least Valued

Table 7

AMTESS Device Feature Ranks Based on the Four Dimensions of Value in the Multiattribute Utilities Analysis (adapted from Criswell et al., 1983, p. 40, 49)

	Dimension of Value			
Grumman Feature	Concept	Implementation	Operation	Motivation
3-D module	1[a]	1	5	4
Student performance record	3	8	6	8
Videodisc	4	6	7	1
Touch panel	7	3	4	2.5
Request help	5	2	1	5
Repeat lesson option	6	5	3	6
Mandatory instructor call after 2 errors	8[b]	7	8	7
Performance feedback	2	4	2	2.5

Seville/Burtek Feature	Concept	Implementation	Operation	Motivation
3-D module	2	1	2	1
Student performance record	1[a]	2	1	5
Slide projector unit	7	6	9	8
Student response panel	9[c]	8	4	9
Student CRT	7	7	3	7
Remove/replace capability	5	3	8	2
Random malfunction selection	7	5	6.5	6
Performance feedback	3	9	6.5	4
Troubleshoot only mode	4	4	5	3

[a]1 = most valued

[b]8 = least valued

[c]9 = least valued

Table 8

Summary of Positive and Negative Statements about the AMTESS Devices (adapted from Unger et al., 1984c)

Grumman Device	
Positive:	Respondents hold favorable opinions of the device. Use of high technology microelectronics and video storage. The device is safer than operational equipment. Automated features (request help, pre-lesson check, feedback, malfunction insertion) are valuable. The ability to perform troubleshooting tasks on the 3-D module is a valuable feature. The videodisc system is an effective motivating feature. Students liked features including the 3-D module, the request help feature, the videodisc system, proceduralized lessons, and lessons addressing STE/ICE.
Negative:	The device frequently malfunctions. Lessons are difficult to edit. Some lessons are too simple or inappropriate. The student performance record is of little value. System response time is too slow. Rebooting is a poor method for restarting a lesson. The instructor CRT provides little valuable information. Malfunction insertion is time consuming.
Seville/Burtek Device	
Positive:	Respondents hold favorable opinions of the simulator. Ease of inserting malfunctions is valuable. Performance monitoring is valuable. The simulator is safer than operational equipment. Respondents liked features including feedback to students, training in procedures, and the flexibility of the slide projector unit. "Hands-on" troubleshooting is a highly valued device feature.
Negative:	Students were confused by the materials to which they must attend. The reliability and durability of the 3-D module should be increased. Normal vibrations are sometimes registered as student error.

For the Seville/Burtek device, positive features included feedback to students and good student performance data, while improvements were suggested in the 3-D module and device repair record.

Feature Comparisons

The Klein et al. (in preparation) results from interviews with device experts complement the results in Criswell et al. (1983) and Unger et al. (1984c). Table 9 presents a summary of the findings of Klein et al. (in preparation) which shows that each device had its strengths and weaknesses. These data are the only qualitative data gathered by asking respondents to compare the devices to each other. Using the device preferences of the judges in Klein et al. (in preparation), it appears that a better AMTESS device could be constructed using the Grumman 3-D module, Grumman student station with videodisc frames, touch panel, and on-line performance feedback, and the Seville/Burtek instructor station with its editing capabilities and Seville/Burtek computer. Klein et al. (in preparation) also point out problems with courseware and supplemental material.

Klein et al. (in preparation) also present suggestions from subject matter experts concerning important features which should be included in the design of future AMTESS devices. These include:

- Low physical fidelity, even to the point of replacing 3-D modules with 2-D computer graphics and videodisc
- Touch panel for student input, located on student CRT
- Detailed student performance monitoring, available on-line on instructor CRT and on hardcopy
- Standard, not proprietary, computer language
- Easy to use editing system
- Simply designed instructor station
- 95% availability criterion
- Thorough consideration given to lesson material which needs to be taught

Courseware Instructional Features Checklist

Table 10 presents the completed courseware features checklist including all three types of features, learning principles, instructor, and optional, but helpful features. The criterion applied for each item in the checklist is simply if the feature is accounted for in the device. No other criteria concerning feature quality, such as ease of implementation, were used. Table 10 shows that both devices include at least one feature, programmed into the computer, related to the basic learning paradigm, reinforcement, punishment, and response generalization, but neglect, to some extent,

Table 9

Summary of Klein et al. (in preparation) AMTESS Device Feature Comparisons

Feature	Preferred Device	Comments
Hardware:		
3-D Module	Grumman	
Student Performance Record	Seville/Burtek	
Instructor CRT	Seville/Burtek	
Student CRT	Grumman	
Visual Display	Grumman	
Computer	Seville/Burtek	
Student Input Panel	Grumman	
Editing	Seville/Burtek	
Request Help	Grumman	
Performance Feedback	Grumman	
Cable Connections	-	Grumman cables too short; Seville/Burtek cables heavy and unprotected
Software:		
Language	Grumman	Seville/Burtek language is proprietary
Instructional material	-	Both need greater range of topics (including use of Technical Manual, troubleshooting concepts) and more complex topics
Instructional Support:		
Instructor training	-	Both need improvements
Access to computer	Seville/Burtek	
Lesson editing	Seville/Burtek	
Instructor handbook	-	Both need improvements
Availability/ Maintainability	Seville/Burtek	Grumman device available only about half the time

Table 10

Courseware Instructional Features Checklist for the AMTESS Devices

Learning Principle Features Programmed into Device	Included in Device	
	Grumman	Seville/Burtek
Basic learning paradigm:		
Present material	Yes	Yes
Orienting stimulus	Yes	No
Student input mechanism	Yes	Yes
Provide consequence contingent on response accuracy	Yes	Yes
Reinforcement:		
Praise messages	Yes	Yes
Pleasant tones	No	No
Other	No	No
	(natural reinforcement available on both)	
Punishment:		
Messages	Yes	Yes
Unpleasant tones	Yes	No
Pause	Yes, locks after 2 mistakes	Yes, locks after 3 mistakes
Other	Yes, audio messages	No
Shaping:		
Gradual increase in difficulty within and across lessons	Not assessed in depth, but yes in gaming exercises	Not assessed in depth, but generally no
Change criteria for reinforcement	No	No
Chaining:		
Increase in number of steps performed before reinforcement is delivered	No	No
Prompting:		
Extra help provided if student makes an error	No	No
Fading:		
Decrease in prompts	Not assessed	Not assessed
Stimulus control:		
Frequent practice of new skills	No	No

CONTINUED

Table 10

Courseware Instructional Features Checklist for the AMTESS Devices, continued

Learning Principle Features Continued	Included in Device	
	Grumman	Seville/Burtek
Stimulus generalization:		
Presenting same instructional material in different ways	Not assessed in depth, but generally no	Not assessed in depth, but generally no
Teaching of principles	Yes	Yes, to some extent
Response generalization:		
Teaching general skills	Yes	Yes
Instructor Features		
Basic learning paradigm:		
Present material	Yes, through ing and universal instructor	Yes, through thumbwheel selection and editing
Provide consequence contingent on response accuracy	No	No
Reinforcement:		
Praise messages	Yes, through editing	Yes, through editing
Pleasant tones	Yes, through editing	No
Punishment:		
Messages	Yes, through editing	Yes, through editing
Unpleasant tones	Yes, through editing	No
Pause	NO	No
Shaping:		
Gradual increase in difficulty within and across lessons	Yes, through extensive, complex editing and universal instructor	Not assessed, but generally no

CONTINUED

Table 10

Courseware Instructional Features Checklist for the AMTESS Devices, continued

Instructor Features Continued	Included in Device: Grumman	Included in Device: Seville/Burtek
Shaping, continued: Change criteria for reinforcement	Yes, through editing	Yes, through editing
Chaining: Increase in number of steps performed before reinforcement is delivered	Yes, through extensive editing	Not assessed in depth, but generally no
Prompting: Extra help provided if student makes an error	Yes, through editing	Yes, through editing
Stimulus control: Frequent practice of new skills	Yes, by repeating same lessons	Yes, by repeating same lessons
Stimulus generalization: Presenting same instructional material in different ways	Yes, through editing	Yes, through editing
Response generalization: Teaching general skills	Yes, through editing	Yes, through editing
Instructor selection of next activity	No	Yes
Instructor control rate of material presentation	No	No
Cue enhancement	No	No

CONTINUED

Table 10

Courseware Instructional Features Checklist for the AMTESS Devices, continued

Optional, but Helpful Features	Included in Device Grumman	Seville/Burtek
Student control rate of material presentation	No	No
Hardcopy printout of performance record (assuming information also available on CRT)	No	No
Storage of student records	Yes	No
Select variables available on performance record	No	No
Select variables (types of input) sensed by computer	No	No
Select calculations performed on student input	No	No
Data summaries available for groups of students	No	No
Student request extra help	Yes	Yes

shaping, chaining, prompting, the need for repeated practice in newly acquired skills, and the need for various presentations of the same new skill. These omissions are no doubt sacrificing training effectiveness to an unknown (probably great) extent.

Table 10 also shows that both devices include editing systems which enable the instructor to create new lessons among other things, but in general, the editing systems are difficult and time consuming to use. As mentioned earlier, if the instructor has to edit extensively to override what the computer presents, one main purpose of an automated training device, to free the instructor, is defeated. Both devices should be studied more in depth with respect to learning principle features; it appears both devices could be improved in this area.

Finally, Table 10 shows neither device includes many optional features, but in fact, too many optional features might overcomplicate the instructor station. However, if an efficient arrangement of controls were developed, the devices might increase their applicability by adding optional features.

Readability Analysis

Table 11 presents the results of the Fog readability analysis conducted on the instructor and student handbooks for both AMTESS devices. As shown in Table 11, the average readability of both Seville/Burtek handbooks is college level; the Grumman handbooks are high school level. For both devices, the readability of the instructor handbook is higher than the student handbook. Because readability indexes are sometimes inflated due to multisyllablic words which are accepted to be readable by the user group, the multisyllablic words in the samples were examined. It appears that the reading grade level of the Grumman instructor manuals is not affected by this to any great degree. However, the Seville/Burtek instructor guide score may be higher than the Grumman score because that guide contains more factual information about the actual equipment and, thus, more technical words. The same may be true for the student handbooks, although the Seville/Burtek score reflects the presence of such words as "nonperformable," "dimensional," "configuration," and "representation." The reading grade levels of all four handbooks should probably be around grade 8 or 9, according to data presented by Sticht (1969) and Chapanis (1976) concerning the reading abilities of servicemen in selected occupational areas.

Student Performance Record Analysis

Table 12 presents an annotated sample from the Grumman student performance record. The record is clearly identified with the student's name and social security number. Information is then given about each instructional segment (segments compose lessons). Information is clearly labeled and formatted. Time elapsed and error scores presented for each segment are useful. However, another performance index such as errors per minute, or correct steps minute (a commonly used and valuable index) would probably be more useful than the one provided, actual elapsed time divided by expected elapsed time. This record could be more descriptive in providing segment names, date, and time of day. A delineation of the correct and incorrect steps taken would improve this record.

Table 11

Readability Analysis of AMTESS Device Instructor and Student Handbooks

	Number of Samples Taken	Mean Reading Level	Standard Deviation
Seville/Burtek Instructor Guide	12	14.4	3.3
Seville/Burtek Student Handbook	6	14.8	2.4
Grumman Instructor Handbook	4	12.4	2.8
Grumman Student Handbook	5	9.2	2.1

Table 13 presents an annotated sample from the Seville/Burtek student performance record. The strength of this record is that it gives the time in session for each step, and for each activity within a step. Unfortunately, a code book is needed to translate the record, thus limiting its usefulness. The format needs improvements in spacing and clarity. The Seville/Burtek device does not register students by name (or other data), so important identifying data are not available on this record. (Seville/Burtek instructors are advised to keep hand-written records of each student's performance.) None of the information on the record is labeled. The record contains irrelevant information such as port numbers. Errors are not clearly marked; the use of "???," "bad order," and "deviation" is not helpful as a way to describe the accuracy of student performance. No summary statistics, such as elapsed time, total errors, and total correct steps are given. The wording of step descriptions is poor ("Off push button to pushed" means push button to off position) and the last line "END OF PRINT" means that there are no more data to be printed concerning that student's performance on the exercise; "end of exercise" or "end of printout" might be clearer wording.

Table 14 presents a summary of the results of the student performance record assessment for both devices.

Device expert opinions consistently echo the theme that operational problems with the hardware have significantly impaired effectiveness of both devices. Critical problems with maintainability were not revealed by the 1472C-based checklist. However, neither device can claim a repair record that is satisfactory for purposes of undergoing short-term experimental evaluation, much less for use in training courses [c.f., Klein et al. (in preparation); Unger et al. (1984b)].

Table 12

Annotated Sample of Grumman Student Performance Report

Sample	Annotation
AMTESS Student Report 0/1/0 0:1:54:86	Computer codes (irrelevant to report)
DENT, STUART 123-45-1234	Student Name Social Security Number (file name)
SEGMENT 0: Elapsed time: 59 seconds Number of errors: 0	Segment Number Time elapsed in segment Number errors made during the segment
Performance Index: 0.98	Elapsed time ÷ expected elapsed time
SEGMENT 1: . . . and so on	

Summary of Results

Hardware. Information concerning AMTESS device hardware comes from the results of several components in the analysis: the 1472C-based checklist, performance record analysis, the APS, and from summaries of opinion surveys previously administered. As suggested by the 1472C-based checklist, the designs of both devices incorporated elements critical to performance of device training function. According to checklist results, however, each device fell short in one critical area, generation of student performance records in the event of power loss. Major improvements in this regard are necessary for the devices to become fully effective devices. The performance record analysis and opinion survey data add to the problems with student performance records of both devices. The checklist also revealed a number of improvements that could be made on both devices, but these changes are not critically needed. For example, the Grumman device gives somewhat sluggish feedback and the Seville/Burtek needs improved labels.

The APS results address details of the displays used by the two devices. The results reveal that both the Grumman display of primary instructional material and the Seville/Burtek display of ancillary material are consistently formatted. However, both displays tend to contain too much information on any one presentation to be easily understood and used by an observer.

Table 13

Annotated Sample of Seville/Burtek Student Performance Report

Sample	Annotation
11 MO FILAMENT TEST -------------------------------	Exercise Number and Name
PA Beam CB...should be off T Control is .402...should be .45	3D switch setting requirements
00:00:00 Step 1 00:00:28 Step 2 ⋮	Time in session; step number; Blank space after step number means correct completion
00:04:38 Step 4	
00:04:42 Port 22 Power On ⋮	Time in session; Port 22 is a computer code location (irrelevant to report); student turned power on; blank space after "power on" means correct completion.
00:05:46 Step 6	
00:07:28 Port 41 Off Push Button to Pushed... Bad Order 00:07:29 Port 41 Off Push Button to Released Bad Order	Student pushed button in, then released it. Bad Order means it should have been released, then pushed in; but this is not enough of a problem to be called an error.
00:07:30 Port 22 Min Power Sw to Off...Deviation ⋮	Deviation is an error; switch should not have been turned off.
00:10:14 TX to Radiate Ready ⋮	Last step completed at time 10 mins 14 sec into lesson; student has pressed button on transmitter (TX); action caused "radiate ready" light to illuminate; blank space after Radiate Ready means action done correctly.
MO tuning Control is .401 should be .450	3-D switch setting requirements at end of exercise.
End of Print	Exercise output completed.

Table 14

Assessment of Student Performance Records for Both AMTESS Devices

Feature	Present on Record	
	Grumman Device	Seville/Burtek Device
Student identification	Yes	No
Date	No	No
Time of day	No	No
Lesson step name	No	Yes, but coded
Time elapsed in lesson/ step	Yes	No, but could be extrapolated
Number errors in lesson/ step	Yes	No, but could be extrapolated
Other performance score	Yes	No
Items labeled	Yes	No
Understandable English	Yes	No
Clear format	Yes	No
Summary data available across lessons/steps	No	No
Charted data available	No	No
Summary data available across students	No	No

Software. Ease of use and flexibility of software are addressed by the 1472C-based checklist and by summaries of the previously collected opinion survey data. The 1472C-based checklist revealed some needed software changes. The most important changes needed concern cumbersome editing packages. Other changes suggested were minor. For example, the Grumman device should present a message explaining delays in processing (e.g., "I'm working now"), and the Seville/Burtek needs prompts which are more clearly worded.

Courseware Features. The instructional features checklist provided the most information in this area, and was supplemented by the opinion survey data and the performance record critique. Use of the checklist revealed the information that both devices present instructional sequences in accordance with only few learning principles. Thus, it is likely that the courseware of both devices could be substantially improved in this regard. As far as the courseware features which were included, such as performance feedback and the request help feature, opinion was generally favorable.

The analysis of student performance records coupled with opinion survey results indicates that both records need improvement. Strong points of the Grumman record include student identification data (made possible by the device's strong capability to store individual data), clear format with understandable English, and calculation of summary statistics. Device experts consistently stated, however, that the record would have been more helpful had it listed each step performed by the student, not just given totals for segments (groups of steps). The printout then was not available on-line during the session, but was available only at the end of a segment. The Seville/Burtek student record, on the other hand, was available on-line, step by step. However, the Seville/Burtek record was poorly formatted, not intelligible without a code book, did not provide summary statistics, and did not identify the student (because the device did not create and store student data files.)

DISCUSSION

The prototype devices have been subjected to several evaluations. Unger et al. (1984a, b, c) completed the transfer-of-training effectiveness evaluation and collected a large amount of qualitative (opinion) data concerning strengths and weaknesses of each simulator. Criswell et al. (1983) and Klein et al. (in preparation) have also collected opinion data on device strengths and weaknesses. Data collected by all three sets of investigators are similar in range (similar sets of about 10 general, obvious features are used) and depth (include only overview issues such as "is the feature useful or not?"). Likes and dislikes, as expressed by government representatives, contractor representatives, instructors, and students, appear to be surprisingly common. The themes echoed in the three previous qualitative evaluations are that the devices work, the devices contain different sets of features from one another, neither device is better than the other, and that device disrepair has created enormous problems. Thus, for issues related to a human factors analysis, only these general qualitative findings were available. The present analysis sought to address different issues and look at the devices in greater detail than had been done previously.

The main concerns of this analysis were hardware, software, and instructional features as they relate to seven topics: clear presentation of instruction, student-device interaction, courseware, editing, displays, safety, and maintainability. Analysis components were, therefore, selected on the basis of their pertinence to those three concerns. However, it was found that most components related to more than one concern, and that the categories, hardware, software, and instructional features, have overlapping elements. The topic of display characteristics, for example, relates to hardware (e.g., size of CRT), software (e.g., is the display understandable, easy to use), and courseware features (e.g., does the display clearly present stimulus items, can the instructor change the display as needed to prompt the student). So, while it seems reasonable to address hardware, software, and instructional features, it was found that the issues could not be addressed in neat packages. (The seven topics mentioned above were easier to address; see the conclusions section of this report.)

With respect to hardware, MIL-STD-1472C contained some relevant sections, although hundreds of criteria were not applicable. Some sections of the checklist that we developed were useful in assessing hardware, and our checklist could become the basis of a human factors checklist for hardware in training devices.

Software was also addressed by our 1472C checklist, and as with hardware, certain criteria used in our checklist might be pulled together into software categories and form the basis for a human factors software assessment for training devices.

Hardware and software issues, in whatever form or organizational scheme, are much more frequently addressed than issues related to instructional design of training device courseware. Our coursewar assessment was conducted on a general level (compared to how detailed an instructional feature/design assessment could be), yet this type of assessment appears to be seldom, if ever, accomplished in the area of training devices. The addition of a sophisticated instructional design assessment for training device courseware would greatly improve assessments in this area, and at the same time might serve as a guide for device designers. Improved training devices should ultimately result.

Recommendations for future device design and human factors analyses are presented at the end of this report.

CONCLUSIONS

Data collected during the present analysis address all the questions the study sought to answer.

1. Does the Device Allow for Clear Presentation of Instructional Material?

Generally yes, but both could be improved in this regard. The Grumman device presents full-color text, still, and moving picture frames using a videodisc system. Information displayed is precise, relevant, and immediately usable. The text frames appear to be consistently formatted, but tend to contain too much volume per frame. According to device user opinions, still and moving pictures heighten student interest and present important information clearly. The device clearly signals the student concerning what to respond to in the instructional sequence. Supplemental materials, the instructor and student handbooks, should be revised to make them easier to read.

The Seville/Burtek device presents computer-generated text frames on a CRT, and presents 35mm slides of schematics and full-color art on a separate screen. The CRT frames were not examined because the device was not operational, a limit on this analysis. The 35mm slides were found to be consistently formatted, but need improvement in legibility and a reduction in the volume of information presented per slide. The text material presented is considered by users to be less interesting than material presented by the Grumman device, and students have been confused about what material to attend to during the instruction. The instructor and student handbooks should be revised to make them easier to read.

2. How Well Do the Devices and Students Respond to Each Other?

Generally satisfactorily, but improvements could be made. The Grumman device provides a touch screen for student input which had high interest value, but was annoying to use because the pressure and direction of the touch required was too specific. The request help feature on the touch screen made it easy for a student to review, and it was seen as a strong point of the device. Generally, feedback given was correct, but sometimes the devices registered errors unfairly. Feedback was clearly presented. Sometimes the Grumman device is sluggish to respond, and it is suggested that if response time cannot be shortened, the device should present a status report (e.g., "Please wait on me.") to the user. The Grumman device has the very strong capability to store student performance data in files. However, users felt that more varied performance data should be collected, and collected more frequently during the lesson. Users also thought it would be helpful to have on-line access to those data. Presently, only off-line access via the hardcopy printer is available. The record was consistently formatted and easy to understand.

As for the Seville/Burtek device, the student enters responses on a push button panel. While the push button labels were found to be unclear, the action of the panel was considered by users to be satisfactory. More detailed information about responsiveness of the device was not gathered because the device was not operational during the analysis. Many users found the 3-D module to be overly sensitive to student manipulations, and often caused the device to score errors when errors were not made. The response time of the device was found to be satisfactory. In contrast to the capability of the Grumman device, the device gathered student performance data on-line, sampled every step taken by the student, reported these data on-line on the instructor CRT and on hardcopy, but did not store student data in files so off-line data retrieval is not possible. This was seen as a limitation of the device. Interpreting the student performance record required use of a code list, and the record, therefore, needs to be redesigned.

3. Is Device Courseware, or Instruction, Based on Learning Principles?

Neither device's courseware appears to be based on important learning principles. Ultimately, of course, determining beneficial changes in instructional sequences is an empirical matter, but significant student performance improvements could be predicted if the courseware were modified to reflect learning principles. Noticeably absent from the present courseware is concern for proper sequencing, introducing new material in order of difficulty, mastery of new material before more difficult material is presented, increase in the amount of practice required, and review or retest of previously mastered material. These courseware improvements are suggested on the basis of only a relatively cursory assessment of instructional features included in the present analysis; a more complex assessment would probably be even more instructive. Both devices present instruction based on a gross view of instructional theory: they present material, require a student response, and provide feedback. A wealth of knowledge about instructional design is available which may be incorporated into courseware improvements..

4. Do the Devices Allow the Instructor to Change the Program?

Yes, for both devices, but for the Grumman device, editing appears to be more unwieldly than for the Seville/Burtek device for two reasons. One, the course material on the Grumman device is presented by videodisc, and only a few companies in the country have the capability to produce or modify videodiscs. Editing Seville/Burtek material may be accomplished by the instructor by changing a computer program and/or by replacing 35mm slides. Two, it is within the capability of an instructor using a Grumman device to rearrange lesson sequences and change feedback messages presented to students, but it is a difficult programming task. An instructor selects a lesson sequence on the Seville/Burtek device before <u>every</u> lesson by simply dialing in the code number of the sequences to be presented. Within-sequence changes and modification to feedback messages may also be made more easily than similar changes may be made on the Grumman device.

5. Are Displays, Labels, Signals, and Controls Easy to Use or Understand?

For critical displays, yes for both devices; for less than critical items, some changes may be suggested. Users have suggested that the Grumman device employ different signals to indicate "pay attention to the CRT" and "you have made an error" rather than have the same signal serve both functions. In addition, the durability of Grumman labels could be improved.

For the Seville/Burtek device, signals should be added that differentially call the student's attention to the presentation of material and feedback. In addition, labels on the student panel are ambiguous.

6. Are Personal Safety Hazards Associated with the Device?

Both devices provide the advantage, over training on actual equipment, of providing danger-free practice in maintenance tasks. No hazards were reported to be associated with use of the Seville/Burtek device. A drawer on the instructor station of the Grumman device, however, which is normally extended, has shin-high squared rather than rounded corners which have created a safety hazard and caused leg injuries (cuts, bruises).

7. Are the Devices Usually in Good Working Order and Easy to Repair?

Neither device can claim an availability record that is acceptable for purposes of extended device testing or use in training. Both devices should be more rugged. As for repairs, the Seville/Burtek device did not incorporate a modular design which affords easy access to device components. In addition, that device employed cryptic labels on some components which might confuse repair personnel.

RECOMMENDATIONS

Future Training Device Design

The most important recommendation arising from the human factors analysis of the AMTESS devices is that courseware design must be emphasized much more than it was during design of the AMTESS devices. Principles of learning and instructional design should be viewed as the primary source of ideas for hardware and software device features. Certainly, courseware and hardware must fit together, but courseware should not be designed around hardware while hardware is left free to develop independently. New avenues of research are suggested by this recommendation because determining effects of different instructional procedures and sequences in courseware is an empirical matter. Studies in this area might include the effects of varying levels of practice on acquisition and retention, the effects of different chaining procedures, and the effects of systematic review on student performance. Retention has been found to be generally poorer in computer-based than in conventional textbook courses (Splittgerber, 1979). Studies of this nature will suggest capabilities that should be automatically performed by the computer.

A second recommendation concerns a reassessment of which functions are automatically programmed into the computer and which functions are allocated to the instructor and student. The device should automatically perform training according to learning principles so the instructor is free to provide extra help. At the same time, it has been suggested that if a student can control the rate of material presentation, he or she may proceed more quickly than if the device or instructor control rate (Splittgerber, 1979). Again, empirical investigations are suggested. Instructor editing procedures should be simplified and thoroughly documented.

The place of the device in the total training course should be more carefully assessed in the future. In the case of the AMTESS devices, neither courseware package fit into the courses for which they were designed. The role of the instructor must be carefully considered, and decisions made about how the instructor, device, and supplementary materials are to be combined. It may be the case that the computer can perform more functions than it does presently, and obviate the need for handbooks. The supplementary materials provided for the Grumman AMTESS device were course introductions, but much technical information was included in the Seville/Burtek handbooks. Who does what, when during the total training course must be carefully considered.

More careful attention should be given to displays and to the composition of frames. The displays on both AMTESS devices tended to be too cluttered. The user should not have to attend to two different things at once. The usefulness of still and moving pictures as text replacements or adjuncts should be studied.

Finally, more attention should be paid to the use of student performance data. Student performance data should be used to evaluate individual student progress, and to assess the effectiveness of the presentation material. Storage of individual files and data summaries should be important in course design.

Future Human Factors Analyses

As mentioned previously, Smode's (1972) broad definition of a human factors analysis includes analysis of the type conducted in the present study in addition to empirical training effectiveness studies. For purposes of this report, recommendations about empirical training effectiveness studies are omitted. Two general topics described below are related to this type of human factors analysis: when it is to be conducted, and what it should contain. First, as pointed out in Figure 4, a human factors analysis should be conducted as part of the FEA, and acceptance criteria at all three stages of device development should include human factors criteria.

Second, it seems important to continue to include several of the analyses used in the present study, and to consider some additional, new assessments. The MIL-STD-1472C-based checklist was an important source of information about hardware and software. If it is used again, it should be reviewed to omit many items which were found to be not applicable. It may be desirable to combine related items, especially less critical ones, into one item, and to increase the number of items related to critical concerns. It may be helpful to abandon the 1472C outline, and list items on the checklist in descending order of criticality, which might simplify analysis of the data. Smode's (1972) instructor station criteria checklist, now difficult to use and apply because criteria are referred to only by a code number, might be reviewed to determine if it provides information which would augment the 1472C-based checklist.

The courseware assessment with its three categories of instructional features proved to be a very important component of the analysis. Its assessment of learning principles is its critical part. The checklist could be expanded to include instructional procedures shown to be effective under certain conditions, such as use of mnemonics and imagery as a memory aid. Full application of the checklist requires that the device be operational and this should be required of future analyses.

Display analysis is also an important component, and should be retained and expanded. The APS used in the present study addressed seven general qualities of a display, and was difficult to use because its jargon is hard to understand. Display analysis should be expanded to include criteria on format, paging, ease of use, and pictures based on recent work in the area of display (Jonassen, 1982). Readability might be enhanced under display analysis.

Two new assessments may be recommended: objective evaluations of editing capability, and of how easily student and device respond to each other. Both these topics are important in assessing a device from a human factors perspective, and were not addressed by the present analysis as thoroughly as might be desirable.

Opinion surveys should continue to occupy the minor role they played in the present analysis. If used, however, it is recommended that surveys be short and ask questions pertinent to the specific areas of professional expertise of the subject. For example, instructors not trained in course design or learning principles should not be asked questions about those areas. It is important to retain user opinion surveys on the level of "do you like the device and why?" (at a minimum) because user comments can influence design, and because user acceptance of computerized training devices is critical to their potential widespread use.

REFERENCES

Air Force Systems Command. (1969). Personnel subsystems. In AFSC design handbook 1-3. Washington, DC: Andrews Air Force Base.

Applied Psychological Services, Inc. (1970). Manual of instructions for the analytic profile system. (Contract N00014-66-C0183 NR 196-076). Wayne, PA: Author. Sponsor: Engineering Psychology Programs Office of Naval Research, Arlington, VA.

Bailey, J. S., & Hughes, R. G. (1980). Applied behavior analysis in flying training research. Williams Air Force Base, Arizona: Flying Training Division, Air Force Human Resources Laboratory. Sponsor: HQ Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, TX. (AFHRL-TR-79-38).

Brock, J. F. (1978). Maintenance training and simulation: design processes and evaluation criteria. Proceedings of the 22nd Annual Meeting of the Human Factors Society, p. 260-266.

Caro, P. W. (1977). Some factors influencing Air Force simulator training effectiveness (Tech. Rep. 77-2). Alexandria, VA: Human Resources Research Organization. Sponsor: Air Force Office of Scientific Research (NL), Bolling Air Force Base, Washington, DC.

Catania, A. C. (1979). Learning. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Chapanis, A. (1976). Engineering psychology. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally.

Cohen, V. B. (1983). Criteria for the evaluation of microcomputer courseware. Educational Technology, 23(1), 9-14.

Criswell, E. L., Unger, K. W., Swezey, R. W., & Hays, R. T. (1983). Army maintenance training and evaluation simulation system (AMTESS) device development and features. (Science Applications, Inc. Report No. 83-03-178) (ARI Tech. Report 589). Alexandria, VA: U.S. Army Research Institute.

Davis, E. G., & Swezey, R. W. (1981). Human factors considerations in the National Training Center 1: Preliminary guidelines for core instrumentation subsystem (CIS) development. (SAI Report No. SAI-81-030-178). McLean, VA: Science Applications, Inc.

Davis, E. G., & Swezey, R. W. (1983). Human factors guidelines in computer graphics: A case study. International Journal of Man-Machine Studies, 18, 113-133.

Faust, D. G., Swezey, R. W., & Unger, K. W. (1984). Field application of four models designed to predict training device transfer-of-training potential. (SAI Report No. SAI-84-06-178). McLean, VA: Science Applications, Inc.

Fink, C. D., & Shriver, E. L. (1978). Maintenance training simulators at Air Force technical training centers: Present and potential use. Alexandria, VA: Kinton, Incorporated. Sponsor: Technical Training Division, Air Force Human Resources Laboratory, Lowry Air Force Base, CO. (AFHRL-TR-78-77).

Garlichs, E. A., Miller, E. J., & Davis, E. (1983). Army maintenance training and evaluation simulation system (AMTESS): Trainer engineering design report. Volume I, chapters 1-6. (Seville Technical Report TR83-14). Pensacola, FL: Seville Research Corporation. Sponsor: U.S. Army, Program Manager for Training Devices (PM TRADE), Naval Training Equipment Center, Orlando, FL.

Hendricks, D. E., Kilduff, P. W., Brooks, P., Marshak, R., & Doyle, B. (1983). Human engineering guidelines for management information systems. Aberdeen Proving Ground, MD, U.S. Army Human Engineering Laboratory.

Hritz, R. J., Harris, H. J., Smith, J. A., & Purifoy, G. R. (1980). Maintenance training simulator design and acquisition -- handbook of ISD procedures for design and documentation. (Tech. Rep. AFHRL-TR-80-23). Brooks Air Force Base: AFHRL.

Jonassen, D. H. (1982). The technology of text. Englewood Cliffs, NJ: Educational Technology Publications.

Kane, J. J., & Holman, G. L. (1982). Training device development: Training effectiveness in the Army system acquisition process. (SAI Report No. 82-02-178). McLean, VA: Science Applications, Inc. Sponsor: U.S. Army Research Institute, Alexandria, VA.

Klein, G. A., Gordon, S. E., & Palmisano, M. (in preparation). Comparison-based predictions and recommendations for Army maintenance training devices. Yellow Springs, OH: Klein Associates. Sponsor: Army Research Institute, Alexandria, VA.

Malone, T. B., Kirkpatrick, M., Mallory, K., Eike, D., Johnson, J. H., & Walker, R. W. (1980). Human factors evaluation of control room design and operation performance at Three Mile Island - 2, Vol. 2. Final Report. Washington, DC: U.S. Nuclear Regulatory Commission.

MIL-STD-1472C. (1981). Human engineering design criteria for military systems, equipment, and facilities. Washington, DC.

Orlansky, J., & String, J. (1981). Cost-effectiveness of maintenance simulators for military training. (IDA Paper P-1568). Arlington, VA: Institute for Defense Analyses. Sponsor: Deputy Under Secretary of Defense for Research and Engineering, Washington, DC.

Powers, R. B., & Osborne, J. G. (1975). Fundamentals of behavior. St. Paul: West Publishing Company.

Ramsey, H. R., & Atwood, M. E. (1979). Human factors in computer systems: A review of the literature. (SAI Report No. SAI-79-111-DEN). Englewood, CO: Science Applications, Inc.

Skinner, B. F. (1968). The technology of teaching. New York: Appleton-Century-Crofts.

Smode, A. (1972). Training device design: Human factors requirements in the technical approach. (Tech. Rep. NAVTRAEQUIPCEN 71-C-0013-1). Orlando, FL: Naval Training Equipment Center.

Splittgerber, F. L. (1979). Computer-based instruction: A revolution in the making? Educational Technology, 19(1), 20-26.

Sticht, T. G. (1969). Literacy demands of publications in selected military occupational specialties. Draft Professional Paper, Human Resources Research Note, Division 3 (Recruit Training), George Washington University, Washington, DC.

Swezey, R. W. (1979). An application of a multi-attribute utilities model to training analysis. Human Factors, 21, 183-189.

Systems Exploration, Inc. (1983). Man-computer functional requirements (NOSC Contract N00123-80-D-0228 Task 7N4Y). San Diego, CA: Author.

Unger, K. W., Swezey, R. W., Hays, R. T., & Mirabella, A. (1984a). Army maintenance training and evaluation simulation system (AMTESS) device evaluation: Volume III, qualitative assessment of two prototype devices. SAI Report No. SAI-84-07-178. McLean, VA: Science Applications, Inc.

Unger, K. W., Swezey, R. W., Hays, R. T., & Mirabella, A. (1984b). Army maintenance training and evaluation simulation system (AMTESS) device evaluation: Volume II, transfer-of-training assessment of two prototype devices. SAI Report No. SAI-84-05-178. McLean, VA: Science Applications, Inc.

Unger, K. W., Swezey, R. W., Hays, R. T., & Mirabella, A. (1984c). Army maintenance training and evaluation simulation system (AMTESS) device evaluation: Volume I, overview of the study effort. SAI Report No. SAI-84-10-178. McLean, VA: Science Applications, Inc.

U.S. Army, Material Development and Readiness Command. (1984). Human factors engineering during system development. (Draft, revised). Alexandria, VA: DARCOM-R.

Van Cott, H., & Kinkade, R. (Eds.) (1972). Human engineering guide to euipment design. Washington, DC: American Institute for Research.

Woelfel, J. C., Duffy, P. J., Unger, K. W., Swezey, R. W., Hays, R. T., & Mirabella, A. (1984). A critical review of front-end analysis procedures employed in the Army maintenance training and evaluation simulation system (AMTESS). SAI Report No. SAI-84-01-178. McLean, VA: Science Applications, Inc.