

1-1-1990

ASTAR Operational Evaluation Final Report Technical Documentation Volume I: Project Description

Merrell M. Bailey

Robert Bird

Brenda A. Bradley

Michael A. Companion

Jorge Franchi

See next page for additional authors

Find similar works at: <https://stars.library.ucf.edu/istlibrary>
University of Central Florida Libraries <http://library.ucf.edu>

This Research Report is brought to you for free and open access by the Digital Collections at STARS. It has been accepted for inclusion in Institute for Simulation and Training by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

Recommended Citation

Bailey, Merrell M.; Bird, Robert; Bradley, Brenda A.; Companion, Michael A.; Franchi, Jorge; and Gibbons, Stephen, "ASTAR Operational Evaluation Final Report Technical Documentation Volume I: Project Description" (1990). *Institute for Simulation and Training*. 25.
<https://stars.library.ucf.edu/istlibrary/25>

Creator

Merrell M. Bailey, Robert Bird, Brenda A. Bradley, Michael A. Companion, Jorge Franchi, and Stephen Gibbons

INSTITUTE FOR SIMULATION AND TRAINING

ASTAR OPERATIONAL EVALUATION

FINAL REPORT
TECHNICAL DOCUMENTATION

Volume I: Project Description

Final Report
CDRL A003

April 13, 1990

(Revised May 11, 1990)

Prepared under Contract Number 61339-89-C-0029

for

Naval Training Systems Center

IST

Institute for Simulation and Training
12424 Research Parkway, Suite 300
Orlando FL 32826

University of Central Florida
Division of Sponsored Research

G 01



ASTAR OPERATIONAL EVALUATION
FINAL REPORT
VOLUME I: PROJECT DESCRIPTION

Final Report
CDRL A003

Michael A. Companion

April 13, 1990
[Revised May 11, 1990]

Prepared under Contract Number 61339-89-C-0029
for the Naval Training Systems Center

by

University of Central Florida
Institute for Simulation and Training
12424 Research Parkway, Suite 300
Orlando, Florida 32826

APPROVED BY:


Michael A. Companion, Ph.D.
Principal Investigator

APPROVED BY:


Daniel Mullally
Program Manager

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGEMENTS

Appreciation is expressed to the following persons for their contributions to this study:

Maj. Wes Woodruff, USAF, NAVTRASYSCEN Liaison Office, Orlando, FL, Dr. Michael Singer, ARI, Field Office, Orlando, FL, and Mr. Tim Whitten, NAVTRASYSCEN, Orlando, FL, for their participation in the Project Advisory Group.

Mr. Russell Irvine, NAVTRASYSCEN, Orlando, FL, Mr. Richard D. Leavitt, Computer Science Corporation, Orlando, FL, Captain Michael Pontius, US Army PM TRADE, Orlando, FL, for their participation in Operational Test #1 and the Longitudinal Test.

Dr. Dennis C. Hribar, Mr. Phillip L. Stuckemeyer, and Mr. D.L. Stokes, Newport News Shipbuilding Co., Newport News, VA, for their participation in Operational Test #2.

Dr. John Lett, Mr. Al Scott and Ms. Sabine Atwell, DLIFLC, Monterey, CA, for their participation in Operational Test #3.

FOREWORD

The research report herein was conducted for the United States Navy by the University of Central Florida, Institute for Simulation and Training under Contract Number 61339-89-C-0029. The Naval Training Systems Center, Orlando, Florida, administered the contract which was sponsored by the Joint Service Committee on Manpower and Training Technology. The contract was to perform an ASTAR Phase III Test and Evaluation, and to produce a Final Report and Transition Plan. The research reported on in this report was performed between the period May 1989 and April 1990.

The Final Report Technical documentation, a three volume set, covers the one year effort to test and evaluate the Automated Simulator Test and Assessment Routine (ASTAR) under operational conditions. ASTAR was compared to a related Device Effectiveness Technique (DET), the Automated Instructional Media Selection (AIMS) system. **Volume I: Project Description**, contains an overview of the entire evaluation effort which includes summaries of all the tests conducted. **Volume II: Test Reports**, contains a full, comprehensive report of all Operational, Longitudinal and Analytic investigations conducted in support of this research effort. **Volume III: Functional Description**, is a comprehensive system specification of the recommended improvements for a new ASTAR, ASTAR II.

EXECUTIVE SUMMARY

PROBLEM

The Automated Simulator Test and Assessment Routine (ASTAR) is an automated decision aid designed to assist a training system analyst to predict the effectiveness of a training device during its development (Rose, Martin & Wheaton, 1988). ASTAR was developed to provide a systematic and analytic evaluation procedure to aid training device design and acquisition. Prior to implementation as a standard evaluation technique, it was necessary to conduct field tests with operational analysts to determine user acceptance of ASTAR.

OBJECTIVE

The objective was to compare and contrast ASTAR to other automated Device Effectiveness Technologies (DETs) and formulate a plan to implement ASTAR as a standard evaluation technique within the DoD Instructional System Development (ISD) process.

APPROACH

The operational evaluation was accomplished through a series of integrated tests using operational training systems and their analysts. The tests assessed the operational utility and impact of ASTAR on existing and new training systems. A single test could not adequately or efficiently address the scope of the evaluation criteria required to assess the operational utility of ASTAR. Therefore, several tests were conducted during the course of this project, including three operational tests and a longitudinal test. These tests examined performance, utility and user issues with regard to ASTAR.

CONCLUSIONS

While the concept of ASTAR was well received by the operational analysts, the current implementation of ASTAR achieved poor user acceptance. ASTAR requires extensive enhancement before it can gain general user acceptance. A functional description for an improved ASTAR was developed which addressed the problems in ASTAR. It is recommended that any further developmental action on ASTAR be limited to consideration of the improved ASTAR, or a totally new effort to develop a technique for estimating training effectiveness.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

	<u>PAGE</u>
INTRODUCTION	1
Purpose of Task	1
Benefits of ASTAR	1
Description of Evaluation DET	2
ASTAR	2
AIMS	5
BRIEF HISTORY OF KEY EVENTS	7
Bibliography of ASTAR Reports	8
MAJOR PROBLEMS -- LESSONS LEARNED	11
Analyst Continuity	11
Evaluation Organization	11
DET Integration	11
Transparency of ASTAR Formulas	12
Study Approach	12
TEST RESULTS	13
Introduction	13
Operational Test #1	14
MCTFIST	14
GUARD FIST I	14
Method	15
Results	16
Conclusions	18
Operational Test #2	19
SEAWOLF IAL	19
Method	19
Results	20
Conclusions	23
Operational Test #3	23
Intelligence Analyst	24
Method	24
Results	25
Conclusions	27
Longitudinal Test	27
SIMNET	28
Method	28
Results	29
Conclusions	31

TABLE OF CONTENTS

	<u>PAGE</u>
Analysis of Operational Tests	32
Summary of User Comments	34
PAG Assessment	35
TRANSITION/IMPLEMENTATION PLAN	37
Need for DETs	37
ASTAR Support of the Interservice Procedure for Instructional Systems Development (IPISD)	38
Implementation of ASTAR	39
Limitations	39
Configuration Control	39
Transition Materials	39
Training Approaches	40
RECOMMENDATIONS	41
Current ASTAR	41
Recompiled ASTAR	41
ASTAR II	41
Unresolved Issues	41
AIR Recommended Changes	42
BIBLIOGRAPHY	45

LIST OF FIGURES

	<u>PAGE</u>
Figure 1. The ASTAR Model of Training Device Effectiveness	4
Figure 2. AIMS Concept of Operation	6

INTRODUCTION

The ASTAR is an automated decision aid designed to assist a training system analyst to predict the effectiveness of a training device during its development (Rose, Martin & Wheaton, 1988). ASTAR was developed to provide a systematic and analytic evaluation procedure to aid training device design and acquisition. ASTAR is a computer-based decision aid developed by the American Institutes for Research (AIR) under contract to the government. ASTAR has been under development and validation for a number of years. This proposal addresses the final phase in the ASTAR project: the operational evaluation of ASTAR.

As the developers of ASTAR stated:

ASTAR is intended to provide training system designers and developers with various kinds of information about the potential effectiveness of a training-device-based system. ASTAR is not designed to produce a single 'Figure of Merit.' The approach to effectiveness analysis is to provide a framework in which device developers can compare devices for effectiveness and diagnose potential problems in a system design. (Rose, Martin & Wheaton, 1988, p. 6)

ASTAR was derived from an earlier technique labeled DEFT (Device Effectiveness Forecasting Technique). The emphasis of these techniques has been to transform an analyst's information and judgements on a training system into predictions of training device effectiveness. ASTAR uses generally-accepted training principles involving such issues as performance, feedback, and similarity of the trainer to operational equipment to evaluate the potential effectiveness of the training system. ASTAR has several levels of analysis that make it applicable throughout the training equipment acquisition process.

PURPOSE OF TASK

ASTAR has been the object of an extended development and evaluation process. It has been "validated" on a number of training systems, including the Portable Aircrew Trainer (PAT), Precision Gunnery Training System (PGTS) and Combat Talon II (CT-II) maintenance trainer. This project was designed to accomplish the final phase of the ASTAR development process; the conduct of "field testing" in order to demonstrate the operational utility of ASTAR. This was accomplished by applying it to a variety of fielded and emerging training systems. The objective was to compare and contrast ASTAR to other automated Device Effectiveness Technologies (DETs) and formulate a plan to implement ASTAR as a standard evaluation technique within the DoD Instructional System Development (ISD) process.

BENEFITS OF ASTAR

The application of ASTAR in an operational environment is

projected to have a number of significant benefits:

1. It should result in simpler, lower fidelity, and lower cost designs which efficiently meet training objectives.
2. The effectiveness of training systems should be enhanced through the systematic assessment of a device's ability to teach each task.
3. The application of ASTAR should reduce the number of and/or simplify tradeoffs.
4. The development time for a new or modified training system should be greatly reduced.

DESCRIPTION OF EVALUATION DETS

The following descriptions of ASTAR and AIMS provide an overview of the two DETs utilized in this evaluation. The text is taken from the descriptive brochures developed as part of the user survey.

ASTAR

WHAT IS ASTAR?

The Automated Simulator Test and Assessment Routine (ASTAR) is an automated decision aid designed to assist an analyst in evaluating the effectiveness of a training device or method. ASTAR runs on an IBM Personal Computer or compatible computer with dual floppy or single floppy and hard disk drive system, and uses minimal computer storage space.

HOW DOES ASTAR WORK?

ASTAR uses generally accepted training principles, involving such issues as performance, feedback, and similarity of the trainer to operational equipment, to evaluate the effectiveness of any training method that involves practice on job tasks. ASTAR helps the analyst evaluate a training approach by asking questions about the training device features that affect learning difficulty or transfer of training to the job environment. ASTAR converts information and judgements provided by the analyst concerning various facets of the training system into a forecast of the system's effectiveness.

The program has three levels of evaluation based on level of detail available. ASTAR Level 1 uses general ratings from the analyst without building a data base with tasks and subtasks as Level 2 or 3 does. The decision about which level is used depends upon the

amount of information the analyst has about the training device, the operational equipment/performance, the tasks to be trained, and the trainees themselves.

Using the analyst's ratings, ASTAR computes several "effectiveness" scores which can be used to make comparisons among devices or methods. An Acquisition Effectiveness score and a Transfer Effectiveness score provide a basis for comparisons of what is learned on the device and what remains to be learned on the job. These scores can be combined to produce a summary score of Training Effectiveness. Figure 1 indicates the kinds of information ASTAR uses to compute the effectiveness scores.

USES OF ASTAR:

ASTAR has been used successfully in a variety of applications and is intended for use through all of the training system acquisition process. For example, ASTAR has been used to compare the effectiveness of:

- * Using a device-based training system vs. training on the operational equipment, and
- * Two training devices that were being used to train the same tasks.

ASTAR has also been used to:

- * Investigate which of several utilization patterns was most effective for an existing device.

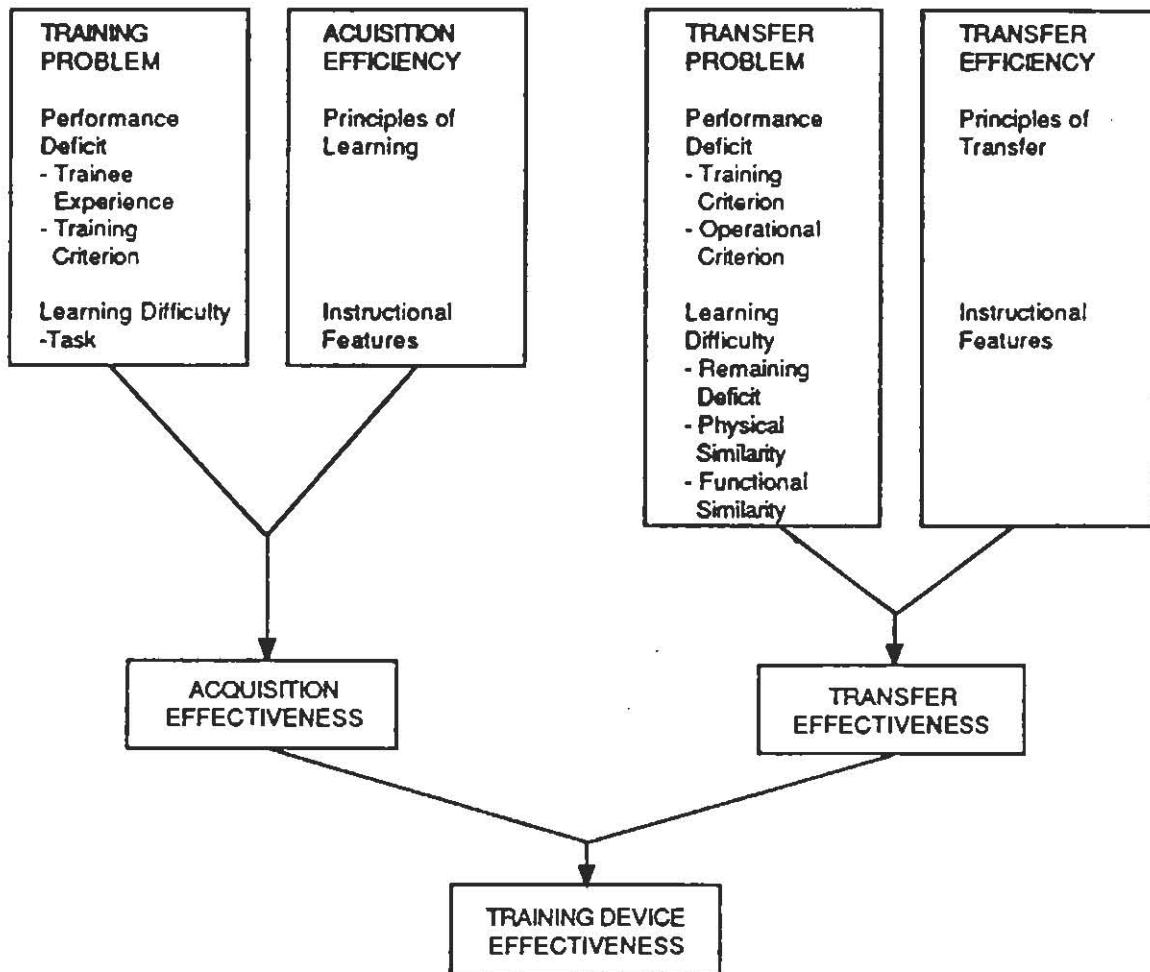


Figure 1. The ASTAR Model of Training Device Effectiveness

AIMS

WHAT IS AIMS?

The Automated Instructional Media Selection (AIMS) System aids the analyst in the selection of media/training equipment to satisfy training requirements. The system is more flexible than other media selection tools because the user can change the definitions and assumptions about media features inherent in the system. AIMS is available to run on an IBM Personal Computer or compatible computer with dual floppy or single floppy and hard drive system and uses minimal computer storage space.

HOW DOES AIMS WORK?

The analyst establishes a set of training objectives and then uses a checklist to identify the media attributes required to train each objective. Media attributes are instructional features, such as types of fidelity, methods of feedback, student-pacing techniques, etc. This information is entered in the computer to allow automatic selection of media which best satisfy the instructional requirements of the objectives. The selected media are ranked in order of appropriateness and the total number of times each medium is selected across all objectives is tabulated and printed out in a report.

Figure 2 represents the AIMS concept of operation. AIMS contains a data base of up to 50 media and 99 media attributes. The analyst can add to or delete from the data base, thereby changing the media model to fit particular needs.

USES OF THE AIMS SYSTEM

The AIMS system can be of use in a variety of analyses, including the following:

- * Determine how training should be divided among training devices and actual equipment.
- * Determine which instructional features of the training device are needed to accomplish training requirements.
- * Compare several training device options to determine what best meets the training requirements.
- * Identify tasks that can be trained more efficiently by using less inexpensive, lower

fidelity devices.

- * Identify which training objectives can be accomplished by using a training device.

Based on use of the AIMS system to date, the system has impacted the acquisition process so as to produce more cost-effective training.

Applications of AIMS have led to the following conclusions:

- 1) It has been applied to emerging aircrew training systems and computer-based training systems.
- 2) It produces information that can be used in cost-effectiveness analyses.
- 3) It can be used by subject matter experts, instructional designers, and training managers.
- 4) It results in better use of simulators, aircraft time, and part-task training.

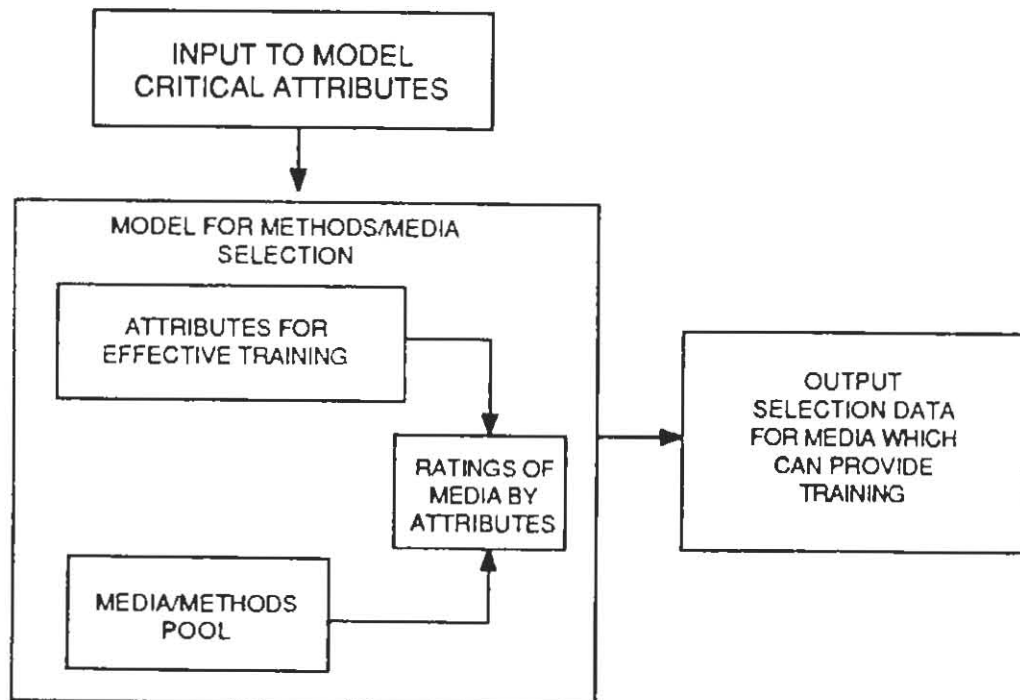


Figure 2. AIMS Concept of Operation

BRIEF HISTORY OF KEY EVENTS

The history of ASTAR began in the late 1970s as a manual analysis technique developed by the American Institutes for Research (AIR) for the U.S. Army Research Institute (ARI). Around 1980, AIR was contracted to convert the manual technique to a computer-based decision aid. The initial version of the program, the Device Effectiveness Forecasting Technique (DEFT), was sponsored by the ARI. Early in the 1980s sponsorship of the program was assumed by the Naval Training Systems Center (NAVTRASYSCEN). Around this time, the name of the program was changed from DEFT to ASTAR.

The ASTAR and DEFT programs were subjected to a series of development and validation tests during the early to mid 1980s. ASTAR was applied to several systems in various stages of the acquisition process to determine its effectiveness as a decision aid. During this research, conducted by AIR under contract to NAVTRASYSCEN, ASTAR was demonstrated to have a positive impact on the design process. ASTAR recommendations influenced the final design selection in the Precision Gunnery Trainer System (PGTS) and Combat Talon II Maintenance Trainer programs. Validity statistics were also established during several tests conducted in conjunction with NAVTRASYSCEN. These validity statistics, summarized by Rose, Martin, and Wheaton (1988) are provided below.

- 1) A split-plot factorial ANOVA indicated that pretraining in Device 11G2 significantly reduced the amount of time to repair (TTR) malfunctions in the Phalanx ($p < .01$) with the effect being greater for more difficult tasks and for certain subsystems (although interaction effects were not statistically significant). Estimated transfer ratio (TR) ranged from .00 to .63 with an average of .33., where

$$TR = \frac{TTR \text{ (in 11G2)} - TTR \text{ (in Phalanx)}}{TTR \text{ (Historical Data)}}$$

(Rose, Martin & Wheaton, 1988, p. 11)

- 2) The ASTAR scale values for each factor were averaged (e.g., scales 1+5/2) to create three variables for predicting transfer ratios. These averages were compared to the empirically derived transfer ratio for each task using a regression analysis ($N = 16$). The resulting multiple correlation ($r = .64$) was described as expressing the goodness of fit between the modeled ASTAR data and the actual performance data.
(Rose, Martin & Wheaton, 1988, p. 12)

- 3) The transfer coefficients developed from ASTAR, FORTE, and field data were correlated two at a

time to provide estimates of concurrent and convergent validity. Convergent validity of ASTAR and FORTE was estimated at $r = .81$ to $.99$ with a mean of $.92$. Concurrent validity of ASTAR prediction of transfer was estimated at $r = .45$ to $.63$ with a mean of $.55$.

(Rose, Martin & Wheaton, 1988, p. 18)

BIBLIOGRAPHY OF ASTAR REPORTS

A number of reports have been developed to document the development and evaluation of ASTAR. The following list identifies major technical reports resulting from the total ASTAR/DEFT project.

1. American Institutes for Research (1987). Review of the DEFT Technique and Recommendations for the Integration of DEFT into the Training Device Design Process. Washington, D.C.: American Institutes for Research.
2. American Institutes for Research (1988). ASTAR User's Manual. Washington, D.C.: American Institutes for Research.
3. Bird, R., Gibbons S., & Companion, M. A. (1990). Operational Test #1: M60A1 Main Battle Tank. Technical Interim Report. Orlando, Florida: University of Central Florida/Institute for Simulation and Training.
4. Bradley, B. (1990). Survey Results. Technical Interim Report. Orlando, Florida: University of Central Florida/Institute for Simulation and Training.
5. Bradley, B. & Companion, M. (1989). The Automated Simulator Test and Assessment (ASTAR) Abbreviated User's Manual. Orlando, Florida: University of Central Florida/Institute for Simulation and Training.
6. Bradley, B. & Companion, M. (1989). ASTAR: The Automated Simulator Test and Assessment Routine. Proceedings of the Cost Effectiveness Analysis Workshop. Held in Conjunction with the 11th Interservice/Industry Training Systems Conference.
7. Bradley, B. & Companion, M. (1990). Operational Test #3: Foreign Language Reading Comprehension. Technical Interim Report. Orlando, Florida: University of Central Florida/Institute for Simulation and Training.
8. Companion, M. & Bailey, M. (1990). Operational Test #2: SEAWOLF Internal Auxiliary Launcher. Technical Interim Report. Orlando, Florida: University of Central Florida/Institute for Simulation and Training.

9. Gibbons, S. & Companion, M. (1990). Longitudinal Test: M60A1 Main Battle Tank. Technical Interim Report. Orlando, Florida: University of Central Florida/Institute for Simulation and Training.
10. Martin, M. F. (1987) Implementation of ASTAR: Evaluation Plan for the Combat Talon II Maintenance Trainer. Washington, D.C.: American Institutes for Research.
11. Martin, M. F. & Rose, A. M. (1988). Implementation of ASTAR: Evaluation of the Portable Aircrew Trainer. Washington, D.C.: American Institutes for Research.
12. Martin, M. F., Rose, A. M., & Wheaton, G. R. (1988) Applications for ASTAR in Training System Acquisitions. Washington, D.C.: American Institutes for Research.
13. Rose, A. M. (1987). Implementation of ASTAR: Evaluation Plan for the Portable Aircraft Trainer (PAT). Washington, D.C.: American Institutes for Research.
14. Rose, A. M. & Martin, M. F. (1988). Forecasting Training Effectiveness: DEFT Final Report. Washington, D.C.: American Institutes for Research.
15. Rose, A. M., & Martin, M. F. (1988). Implementation of ASTAR: Evaluation of the Precision Gunnery Training System. Washington, D.C.: American Institutes for Research.
16. Rose, A. M. & Martin, M. F. (1988). Implementation of ASTAR: Evaluation of the Combat Talon II Maintenance Trainer. Washington, D.C.: American Institutes for Research.
17. Rose, A. M., Martin, A. W., & Yates, L. G. (1985) Forecasting Device Effectiveness: Vol. III. Assessment of Device Effectiveness Forecasting Technique. Washington D.C.: American Institutes for Research.
18. Rose, A. M., Martin, M. F., & Wheaton, G. R. (1988). Forecasting Training System Effectiveness: Review and Analysis of ASTAR. Final Report. Washington, D.C.: American Institutes for Research.
19. Rose, A. M. & Wheaton, G. R. (1984). Forecasting Device Effectiveness: Volume II, Procedures. Washington, D.C.: American Institutes for Research.
20. Rose, A. M., & Wheaton, G. R. & Yates, L. G. (1985). Forecasting Device Effectiveness: Volume I, Issues. Washington, D.C.: American Institutes for Research.

THIS PAGE INTENTIONALLY LEFT BLANK

MAJOR PROBLEMS -- LESSONS LEARNED

During the conduct of this evaluation, a variety of lessons were learned and collected as ASTAR users engaged in the Operational Tests. Observations by the research team as well as the Project Advisory Group (PAG) contributed to the body of recorded lessons learned. These lessons learned are presented below and generally discuss: analyst continuity, evaluation organization, DET integration, transparency of ASTAR formulas, and study approach.

ANALYST CONTINUITY

The tests revealed the importance of using the same raters throughout all phases of an ASTAR analysis and across devices. In one operational test and the longitudinal test, the analyst population was not constant. Even though two analysts participated throughout the analysis process, the impact of the final analyst was significant. ASTAR requires consensus ratings. When the rater population changes, the differences in biases can outweigh the device differences. The investigators had to work with the analysts to statistically factor out the biases introduced by the variable rater population before the devices could be compared. Effective ASTAR analysis requires that the devices be compared and rated from the same point of reference.

EVALUATION ORGANIZATION

Throughout the evaluation, agencies were soliciting for their assistance in conducting on-site evaluation of new training devices. The agencies often promised full support of the effort, but because of project delays, change of personnel, or scheduling conflicts, it was difficult to conduct the operational tests. Numerous agencies and their associated training devices were scheduled to have the analyses conducted, only to be canceled due to one of the above reasons. Since only volunteers were solicited to participate in this evaluation, little or no control over the availability of training systems analysts or training devices was possible. It was necessary to conform to the schedule of the host agencies, while being constrained by our project schedule. The lack of control caused serious delays in initiating and completing the evaluation effort. Tests which are designed to evaluate operational use of a decision aid in the design process need to conform to the schedule of the selected design efforts. In this case the goal was to examine a decision aid that should support the entire three to five year development cycle of a training system, yet the evaluation program was restricted to a single year.

DET INTEGRATION

ASTAR presently exists as a stand alone system which does not have the capability to "talk" to other software decision aids. Analysts indicated that they are being encouraged to use all sorts of automated decision aids. However, most of these

decision aids are stand alone systems. Hence, the analyst is required to enter the same data repetitively. This leads to poor user attitudes and acceptance of decision aids. Users view these "aids" as hindrances which make their job harder. For a decision aid to be accepted, it must be able to "electronically talk" with other decision aids. The ability of ASTAR or other DETs to be integrated is a critical design feature.

TRANSPARENCY OF ASTAR FORMULAS

During the conduct of each operational test, users were almost always confused by what they saw in the summary statistics. This aspect of ASTAR was often criticized. Users simply did not understand the output data, or ASTAR results, as presented in the final summary statistics. The difficulty in determining the meaning of output statistics lies in the fact that the formulas which drive them are not presented to the user either in the program or documentation. It has long been known that users tend to distrust automated devices when they do not know what is going on within the device. This feeling was evident with ASTAR. In order to gain users' confidence in ASTAR, a discussion of the underlying computational formulas should be included in the training materials. The problem is compounded by the fact that the ASTAR formulas are somewhat obscure -- showing them to the analyst may not really clarify how ASTAR works.

STUDY APPROACH

The initial UCF proposal specified the use of training devices previously fielded and in use, with the option to select new systems based on the user survey. This approach facilitated the evaluation of criteria such as cost impact. Based on the survey results, the decision was made to use new and emerging systems for the operational studies to assess the impact of ASTAR on designs. In addition to the problem of schedule discussed in Section 3.3, this decision made it impossible to adequately evaluate several of the target criteria, such as cost impact. With the snapshot approach to the device design imposed by this decision, it was not possible to gather sufficient data to judge cost and development time impact. The total set of criteria identified for the ASTAR evaluation can not be adequately assessed with either approach. A mixture of established and new systems would have been a better choice for the operational studies to provide some degree of assessment for all evaluation criteria.

TEST RESULTS

INTRODUCTION

The operational evaluation of ASTAR was accomplished through integrated research which utilized operational training systems and their analysts. The research assessed the operational utility and impact of ASTAR on existing and new training systems. A single test could not adequately or efficiently address the scope of the evaluation criteria required to assess operational utility of a program such as ASTAR. Therefore, several tests were conducted during the course of this project, including three operational tests and a longitudinal test. These tests examined performance, utility and user issues with regard to ASTAR. In addition, the operational tests compared ASTAR to another automated decision aid, the Automated Instructional Media Selection (AIMS) program, and to conventional methods of training device design.

The three operational tests involved a structured evaluation of ASTAR and AIMS. A minimum of three analysts were used for each training system evaluated. When possible, the subjects selected for research were the actual analysts involved in the original development of the selected training systems. The tests compare ASTAR and AIMS to the conventional approaches used to develop existing training systems. AIMS was used to provide a comparison of ASTAR with another automated decision aid. A user attitude survey was developed for the tests to assess the subjects' reactions to the use of the two techniques. The questions addressed the analysts' acceptance of and attitudes toward the user friendliness and overall usefulness of the decision aids. User attitudes toward the computerized decision aids were evaluated at the completion of each test. Summary results are reported for each decision aid.

A longitudinal test was conducted to provide a test of actual operational use of ASTAR within an ongoing training system development program. The test was originally designed to address the fundamental question of how an analyst would actually use ASTAR during the development of a training system. However, unanticipated delays in program start up precluded the test from using an ongoing development effort for the longitudinal test. Thus, the focus changed to evaluating opinions of analysts who had used ASTAR over an extended period of time. The test was by nature unstructured and emphasized lessons learned. The test tracked two analysts as they applied ASTAR to various training devices over a seven month period. Responses to user attitude questionnaires were solicited after each application of ASTAR. The questionnaires were evaluated for changes over the extended application of ASTAR. Additionally, a termination interview was conducted to assess the analysts' final concerns and opinions regarding ASTAR.

OPERATIONAL TEST #1

The first operational test of ASTAR and AIMS was conducted using the Marine Corps M60A1 main battle tank as the weapon system of interest. The test applied ASTAR and AIMS to the comparative evaluation of the Marine Corps Tank Full-crew Interactive Simulator Trainer (MCTFIST) and the Guard Unit Armory Device Full-crew Interactive Trainer (GUARD FIST I) as potential training devices for the M60A1 tank. A combination of questionnaires and actual results derived from the use of the DETs was used to gather data on ASTAR and AIMS.

This operational test was designed to compare the training effectiveness of MCTFIST and GUARD FIST I on a common subset of the Marine Corps M60A1 tank training objectives. Results should provide insight as to whether GUARD FIST I provides an acceptable training device alternative to MCTFIST. In addition, insight concerning the impact of computer generated imagery (CGI) versus video disk visual scenes on predicted training effectiveness was sought. The primary technology difference was video disk in MCTFIST and CGI in GUARD FIST I.

MCTFIST

MCTFIST is a training system simulates selected gunnery tables, enabling a full tank crew to develop and sustain individual and crew tactical engagement and gunnery skills applicable to the M60A1 main battle tank. The system includes an Instructor/Operator who manages the training and provides comprehensive after-action reviews. MCTFIST is a strap on training device so that training takes place within a stationary, powerless M60A1 tank. All crew members (Tank Commander, Gunner, Driver, and Loader) participate in selected gunnery tasks. The crew observes appropriate visual and aural effects while using actual tank controls to simulate the tank's operation. Training exercises involve simulated cross-country travel and engagements with enemy forces. MCTFIST is transportable and can be rapidly installed wherever desired.

MCTFIST uses CGI to superimpose targets, target signatures, and weapons effects on filmed background scenery to provide a realistic training experience. The CGI allows complete freedom of target placement and movement, while the video disk scenery provides the realism of an actual engagement. The video background reflects varied terrain and provides a ranging and engagement capability from 500 to 2,000 meters. The current MCTFIST scenery was photographed at the National Training Center, Fort Irwin, California and portrays a daylight desert environment.

GUARD FIST I

Like MCTFIST, GUARD FIST I is a full crew trainer that simulates both daytime and thermal engagements. It uses CRTs to present complete CGI scenes. The trainer is mounted on the

Army's M1A1 main battle tank to present targets. These targets can be simulated with either European or desert terrain as background. Other simulated features include: tank movement within a limited area of operation; full 360 degree rotation of the turret; and firing of both the main gun and the coaxial machine gun.

The GUARD FIST I training system provides the means for the M1A1 tank crew to practice full-crew interaction procedures from a stationary tank. Training is conducted with the turret in the travel lock position. Training scenarios present realistic simulated environments that require the crew to respond as they would in combat engagements, using proper full crew interactive procedures, tank controls, and fire control components. Sensors attached to the tank controls provide real-time responses to crew reactions during simulated battle engagement exercises. GUARD FIST I is transportable and can be installed at National Guard Armories and Reserve Centers.

Method

Subjects. Two subjects participated in all phases of the test: a Project Director from NAVTRASYSCEN familiar with training analysis and design, and a contractor representative from the simulator manufacturer, who was familiar with the device and the tank. These subjects were supplemented in the MCTFIST application by a tank gunnery sergeant who served as a subject matter expert (SME). In the GUARD FIST I evaluation, a Project Director from PM TRADE assisted with the analysis. With the exception of the MCTFIST SME, all subjects were highly experienced in the use of computers.

Procedure. ASTAR and AIMS were applied to both MCTFIST and GUARD FIST I.

The MCTFIST application of ASTAR and AIMS was conducted at different times in Tallahassee and Orlando. The portion of the test conducted in Tallahassee provided access to the SME. Following completion of both the ASTAR and AIMS techniques, a debriefing session was held with the subjects to discuss their experiences and opinions, and to identify any problems that would warrant future actions. During the test, subjects were asked to complete an attitude survey designed to assess their reaction to the technology, and to keep a log of several factors concerning evaluation methods and time spent in technique familiarization and actual analysis.

As part of the MCTFIST application, the NAVTRASYSCEN representative was initially given approximately eight hours of training on the operation of ASTAR and AIMS. With a background in training, this subject then developed the media pool and list of the attributes to be used in the AIMS application. He then entered the ratings for the media/attributes matrix. The remaining two subjects were trained on ASTAR and AIMS prior to the conduct of the MCTFIST evaluation.

The GUARD FIST I portion of the test was completed at a later date on a single day. It was conducted in Orlando by the Navy Project Director, the PM TRADE Project Director, and the contractor representative. Because of the participation of two subjects on the MCTFIST evaluation, no additional training was required. The PM TRADE representative served as the GUARD FIST I SME. Following completion of both the ASTAR and AIMS techniques, a debriefing session was held with the subjects to discuss their experiences and opinions.

Results

ASTAR Evaluation. The subjects conducted ASTAR Level 1 and Level 2 evaluations of MCTFIST and GUARD FIST I, comparing them with the operational tank system. The subjects assigned consensus ratings to the ASTAR questions for use as input to the ASTAR system.

ASTAR computes a summary score based on analysts' ratings on eight categories of questions. The scores represent a relative prediction of training effectiveness that is used to compare devices. The lower the score, the higher the predicted training effectiveness. Both the ASTAR Level 1 and ASTAR Level 2 analyses predicted MCTFIST to be more effective at training the Marine tank task requirements identified for this test. The ASTAR Level 1 summary scores were 66.70 for MCTFIST and 105.46 for GUARD FIST I. For the ASTAR Level 2 analysis, the summary scores were 86.71 for GUARD FIST I and 56.88 for MCTFIST.

The ASTAR Level 1 and ASTAR Level 2 scores for the two devices were quite different. An examination of the subscores provided insight into the composition of the difference. Regarding the two basic subscores, acquisition and transfer, most of the difference between the two devices occurred on the transfer portion of the score. This difference in transfer is logical, because for this test, the operational environment was the M60 main battle tank; the MCTFIST is a M60A1 trainer, while the GUARD FIST I is a M1A1 trainer. Since GUARD FIST I was designed for a different tank, there should be a significant transfer problem for the GUARD FIST I. Since performance deficit was the same in both situations, the difference in acquisition reflects the training capabilities of the devices for the M60A1 task environment. Hence, the findings support what would be predicted based on learning/training principles.

AIMS Evaluation. In the AIMS evaluation both devices were selected as acceptable for each of the training objectives. For ten out of the eleven objectives, GUARD FIST I was rated higher than MCTFIST, with a tie on the last objective.

The data base was examined to explain the apparent clear cut advantage for the GUARD FIST I. The examination revealed several factors that might reduce the strength of the finding. The ratings on the two trainers averaged out across most of the

attributes. GUARD FIST I was given a significantly higher rating in only two categories of attributes, communications and physical cues. These two categories of attributes were the primary reason that GUARD FIST I achieved a higher overall rating than MCTFIST. The subjects identified that, on the average, approximately 90% of the attributes were critical. Hence, the specific ratings used in the selection process did not vary significantly from the overall ratings. Furthermore, in the two categories in which GUARD FIST I had a decisive edge over MCTFIST, four of the five attributes were always among the critical attributes. Hence, the selection process was driven by a small number of the total pool of attributes. Generally, a much smaller subset of total attribute pool is identified as critical to a training objective. The findings described suggest that either the number of identified critical attributes was too high or that the total number of attributes was too small.

The results of the AIMS evaluation indicated that the data base was probably inadequate. In addition, the attributes did not adequately reflect the design difference between the two trainers. As stated initially, the GUARD FIST I used computer generated imagery for the visual scene, while the MCTFIST uses video disk supplemented by computer generated cues. The desire was to determine the impact of this technology difference on training, but the attribute pool does not provide sufficient differentiation on this design feature. During any potential implementation of AIMS, it will be critical to teach analysts how to construct/tailor the AIMS data base to address the critical design issues.

Reactions To ASTAR. The overall reaction to ASTAR was quite negative. Of the six items in this category, ASTAR was rated no higher than 2.5 on a scale of 1 = low; 4 = average; 7 = high. ASTAR was perceived as not useful, rigid, unproductive, difficult to use, frustrating, and lacking in power. The responses to questions concerning acceptance of ASTAR indicated that the ASTAR model as it currently exists was not acceptable. Four of the six questions received the lowest possible score.

Perceived positive aspects of ASTAR included:

ASTAR questions were reasonably clear.

ASTAR terminology was relatively consistent and easy to understand.

ASTAR instructional materials were average.

Perceived negative aspects of ASTAR included:

ASTAR screens had illogically organized menus, non-helpful prompts, and confusing labeling.

ASTAR terminology did not keep the user informed of what the program was doing.

During the use of ASTAR tasks couldn't be performed straightforwardly, the feedback and error messages were not helpful, and memory requirements were high.

ASTAR outputs were unusable, difficult to understand, and confusing.

Reactions To AIMS. The overall reaction to AIMS was only slightly below average. Of the six items in this category, AIMS was rated at 3.0 or 4.0 for all items on a scale of 1 = low; 4 = average; 7 = high. Hence, AIMS was perceived as about average on flexibility, usefulness, productivity, ease of use, and power. The response to the questions concerning acceptance of AIMS indicated that AIMS was perceived as slightly above average on all questions.

Although AIMS was given an average rating and achieved a moderate degree of acceptance, most of the detailed ratings were still negative. Only two areas were perceived as average or positive:

AIMS terminology was perceived as consistent and easy to understand.

The instructional materials for AIMS were helpful and thorough.

Conclusions

Overall, the user acceptance of the two techniques was rather low, though AIMS tended to be rated higher than ASTAR. It was evident that ASTAR and AIMS were considered quite worthwhile in concept but somewhat flawed in terms of user friendliness. The user interface was clearly the major factor in determining future acceptance of the two methodologies. This lack of user friendliness overshadowed the potential benefits of ASTAR and AIMS.

The subjects involved in the test approved the concept of automated decision aids to assist instructional developers in the evaluating and comparing training effectiveness in different emerging devices, or of proposed changes to existing devices. For instance, the ability to conduct ASTAR evaluations at three different levels of device development was felt to be of particular benefit. The subjects believed that the proper application of ASTAR should result in considerable savings in time, cost, and man hours during the analysis phases of training development. However, subjects would prefer not to use the programs as they presently exist, because of their unfriendly nature.

The findings of this operational test indicate that both ASTAR and AIMS will require modifications before they are implemented as standard evaluation techniques. Without these

modifications, user acceptance would be poor at best. Most of the shortcomings can be alleviated by modifying the programs to incorporate current software practices, data base techniques, and user interface standards.

OPERATIONAL TEST #2

The second operational test was conducted with the assistance of training analysts from Newport News Shipbuilding, located in Newport News, Virginia. The analysts were asked to use and evaluate both ASTAR and AIMS, and to compare them to the existing methodologies used by Newport News Shipbuilding. The specific subsystem used in the test was the SEAWOLF Internal Auxiliary Launcher (IAL) System. The IAL is an actual subsystem for which training requirements will be finalized during the course of the SEAWOLF program. A combination of questionnaires, self-initiated logs, and actual results derived from the use of the two DETs were used to gather data on ASTAR and AIMS.

SEAWOLF IAL

The function of the Seawolf IAL is to launch both six-inch and three-inch devices for evasion, environmental monitoring, communications or signaling. The IAL system can be operated in a semiautomatic mode, with both tethered and non-tethered device launch capabilities. In addition, the IAL system can be operated with a hand pump, can be manually overridden, or can launch a device with a hand rammer. Differing equipment and training requirements are necessary for each method of operation. The training option used for this test was manual operation by hand pump.

Two alternate training devices being considered were selected for the ASTAR analysis: a 2-D device and a 3-D device. The 2-D device would be a flat mockup representation of the IAL with actuators and indicators. It would have some operable controls and provide feedback to the operator. The 3-D device would be a full-scale three-dimensional mockup of the IAL with operational controls. In addition to the actual shape, depth, and dimensions, the 3-D trainer would provide pressures, sounds, movements, doors, latches, and other elements to replicate the actual operational equipment. The 3-D device would also permit the launch items to be loaded into the training device.

Method

Subjects. Three subjects from Newport News Shipbuilding participated in the test. All three subjects, one training program developer and two ISD subject matter experts, were highly experienced and comfortable using computers.

Procedure. Training on the use of ASTAR and AIMS was conducted on-site at Newport News Shipbuilding over a two-day period. The two DETs were demonstrated to the subjects, with hands-on experience provided through a sample exercise.

Questions were answered as necessary. Identical sample exercise were used to familiarize the subjects with ASTAR and AIMS. The subjects were provided user manuals for both DETs.

The IAL mission critical task, Operate IAL with the Hand Pump, was selected as the task for the test. This task would be performed under casualty/degraded/abnormal modes of operation. It requires twenty steps and six sub-steps. The sub-steps were not used for evaluation purposes. Based on the data, Newport News Shipbuilding, with assistance from IST, developed the ASTAR and AIMS data bases for the test.

The AIMS analysis was conducted as a standard instructional media selection analysis for the IAL task set. The ASTAR analysis compared the 2-D and 3-D IAL training devices. During the test, subjects maintained logs documenting use of the two DETs, including any problems encountered. The User Attitudes Questionnaires were completed at the end of the test. The logs, questionnaires, and resulting ASTAR and AIMS outputs were used as the data for the test.

Results

ASTAR Evaluation. The subjects conducted both ASTAR Level 1 and 2 analyses for the two potential training devices. The training devices were compared to the projected operational Seawolf IAL system. The subjects assigned consensus ratings to each of the ASTAR questions.

Overall, both ASTAR Level 1 and Level 2 evaluations indicated that the 3-D trainer had a better predicted training effectiveness than the 2-D trainer. However, on an absolute basis, the difference in the summary scores for the two training devices was small. Both trainer alternatives achieved apparently good scores based on the ASTAR metrics.

An examination of the ASTAR Level 1 evaluation summaries showed that the summary scores were 26.23 for the 2-D trainer and 22.55 for the 3-D trainer. [Lower scores indicate higher predicted training effectiveness.] Therefore, at an ASTAR Level 1, a 3-D trainer was predicted to be more effective in this situation than a 2-D trainer. The training problem subscore was identical for both devices, which is normal for most situations. This score was relatively low because the trainees were expected to enter the training environment with a small performance deficit, and it was predicted that the learning difficulty of the task was low. The 3-D device was predicted to have both a higher acquisition efficiency and transfer efficiency, as a result of the higher fidelity of the trainer. These two factors were responsible for the overall advantage that ASTAR projected for the 3-D training device over the 2-D training device.

The same overall result was found with ASTAR Level 2. The 2-D training device received an overall score of 25.95, and the 3-D training device received an overall score of 22.31. The summary

scores were slightly lower than the ASTAR Level 1 scores. In this test, the difference in ASTAR Level 1 and Level 2 scores was a function of changes in the training problem subscore. The subscores for acquisition efficiency, transfer problem, and transfer efficiency were the same for both the ASTAR Level 1 and Level 2 analyses.

AIMS Evaluation. The AIMS analysis was conducted with a pool of 25 media and 60 attributes. Of the potential pool of 25 media, only a subset of eight were selected as acceptable media for the Seawolf training objectives. Of the eight, five were selected as possible media for all twenty of the learning objectives, while one was selected for nineteen of the objectives and the remaining two were selected as acceptable for thirteen of the objectives. Overall, a combination of classroom and equipment type trainers were selected as the most acceptable media.

For each objective, AIMS selected the media which met all of the critical attributes. Between five and eight media were selected for each objective. The one anomalous selection was Mediated Interactive Lecture (MIL), which was always selected as the best choice because of the way MIL was defined. MIL was defined as lecture combined with any of the other media, and this combination of two media always had an advantage. After MIL, all but one of AIMS' second choices were either whole task trainer or operational equipment. The only exception was the selection of Interactive Video Disk as the second choice for Objective 1.0.

The 2-D and 3-D mockups used in the ASTAR analysis basically correspond to the mockup and part-task trainers, respectively, in the AIMS media pool. The 2-D configuration used for the ASTAR analysis has slightly higher realism than the definition of mockup in the AIMS media pool. In the AIMS analysis, the part-task trainer was selected for all twenty of the training objectives, while the mockup was selected as acceptable for only thirteen of the Seawolf manual IAL training objectives. The mockup was never rated higher than fifth within the set of selected media and was always rated lower than the part task trainer. Therefore, the AIMS analysis agrees with the ASTAR analysis. The 3-D training device was a better option for training the manual operation of the Seawolf IAL.

Reactions To ASTAR. The analysts' overall reactions to ASTAR were that ASTAR was difficult, frustrating, and unproductive. Ratings were achieved only after much effort. ASTAR was perceived as rather rigid, because after the model was built, the analysts could not edit it without crashing the file. The overall ratings reflect that ASTAR is not a user friendly program. It is relatively old and does not reflect current software practices. Nevertheless, users still feel that there is something worthwhile in ASTAR. If ASTAR was improved to incorporate a better user interface, then it might be accepted by analysts.

The analysts were also asked to discuss their acceptance of ASTAR. They expected to have little use for ASTAR in their daily work. They did not feel comfortable working with ASTAR, and they did not believe that ASTAR would increase their job effectiveness. However, they believed that ASTAR could be used to find solutions to problems that cannot currently be solved. The unwillingness to work with ASTAR, despite its perceived benefit, was driven by the poor user interface in the current version.

Perceived positive aspects of ASTAR included:

The ASTAR screens were average or above average in organization, labeling, prompts, and presentation of questions.

During the use of ASTAR, tasks could be performed in a straightforward manner.

Format of the ASTAR outputs was average in clarity.

Perceived negative aspects of ASTAR included:

Terminology was confusing, inconsistent, and did not keep the user informed of what was going on.

Learning to use ASTAR was difficult, and the instructional materials were not helpful.

ASTAR outputs were not very useful and were extremely difficult to understand.

Reactions To AIMS. Overall, the analysts believed that AIMS was useful and productive. Alternatively, AIMS was considered to have inadequate power due to its lack of a data base and/or spreadsheet software interface for creating and editing the rating matrix. AIMS was also considered rigid for the same reason.

The analysts expressed a somewhat positive acceptance of AIMS. They expected that they could use AIMS occasionally for planning and proposals, and they agreed that they could work with AIMS. They believed that they would feel comfortable working with AIMS, and slightly agreed that knowing how to work with AIMS would increase their job effectiveness. However, they indicated that the services provided by AIMS could be replicated with the use of a spreadsheet, and that it would be very easy to stop using AIMS. AIMS is effectively a specialized spreadsheet as presently implemented, so this was a natural response.

Perceived positive aspects of AIMS included:

AIMS screens had average or above average organization of menus, labels, and helpful prompts.

The terminology in AIMS was consistent, average in understandability, and kept the user informed.

All aspects of learning to use AIMS were average.

The instructional materials were considered well above average.

Perceived negative aspects of AIMS included:

Confusing questions in the AIMS screens.

During the use of AIMS, feedback and error messages were not helpful and memory requirements were high.

AIMS outputs were slightly below average in terms of usefulness and were somewhat difficult to understand.

Conclusions

The Seawolf IAL test case in the operational test provided a good application of the two techniques and permitted the analysts the opportunity to determine whether the two DETs had sufficient merit to be used on other areas of the Seawolf ISD process.

Overall, the user acceptance of the two techniques was rather low. In both cases the users perceived benefits from use of the two programs, but the unfriendly nature of the interfaces made them unacceptable. Users would not use the techniques if given the option. They would rather use their conventional methods. Of the two DETs, AIMS was preferred over ASTAR. However, this appeared attributable to the perception that AIMS was easier to learn and use than ASTAR. The user interface was clearly the major factor in determining user acceptance of the routine. This conclusion was not unexpected. This finding does not imply that it was the only factor influencing user acceptance. In the case of ASTAR, the lack of understanding about how the program internally works, i.e., the formulas and metrics, also caused concern among the subject analysts.

OPERATIONAL TEST #3

The third operational test was conducted with the assistance of training analysts from the Defense Language Institute Foreign Language Center (DLIFLC), located at the Presidio of Monterey, Monterey, California. The analysts were asked to use and evaluate ASTAR and AIMS, and to compare them to existing methodologies used by DLIFLC for the development of language training courses. A combination of questionnaires and actual results derived from the use of the DETs were used to gather data on ASTAR and AIMS.

The DLIFLC training environment and mission are quite different from those examined in other ASTAR and AIMS tests. However, both ASTAR and AIMS should be applicable to any training

environment. Inclusion of DLIFLC in this operational test was intended to provide better insight into the requirements to implement decision aids, such as ASTAR and AIMS, across the DoD training community.

Intelligence Analyst

Many graduates of DLIFLC transition into classified positions. Thus, DLIFLC often does not know the specific tasks performed on the operational job. Instead, DLIFLC knows the general positions filled by graduates and structures its curriculum to teach basic language skills associated with these positions. A common task in their foreign language training is oriented toward intelligence analysts. The trainee in this position is required to speak, read, write and interpret printed and audio information sources. DLIFLC curricula are designed to teach reading and verbal comprehension sufficient to perform the intelligence analyst' job.

For this test, the "Operational System" was defined as the position of an intelligence analyst in a Spanish speaking country. The training environment selected was "level 2," reading comprehension in Spanish. The specification of a language environment was required because DLIFLC personnel indicated that ASTAR ratings would be highly language-dependent.

Two systems for training reading comprehension were chosen to be evaluated and compared during the ASTAR analysis. The two training systems were Computer-Managed Instruction (CMI) and Programmed Text. CMI was defined as a computer-based training system using the Electronic Information Delivery System (EIDS) as a reference. EIDS is an interactive audio-visual training work station that can integrate the delivery of existing audio-visual media such as print, film, videotape, slides, etc. EIDS is currently available within the DLIFLC training device inventory. Programmed Text was defined as a self-paced paper-based text consisting of short training sections followed by test questions that must be correctly answered before continuing.

Method

Subjects. Two subjects from the DLIFLC participated in the test, one specializing in instructional technologies and the other in curriculum development. Based on their responses to the attitude questionnaire, it was evident that the subjects were confident, comfortable, and experienced in using computers on the job.

Procedure. Training on the use of ASTAR and AIMS was conducted on-site at DLIFLC over two half-day and one full-day sessions. The first-half day consisted of brief introductions, descriptions of the purpose and plan of the overall evaluation, familiarization of the subjects with the DETs, and gathering of the initial data for ASTAR and AIMS data bases. The initial data included the definition of the operational environment, the

training task list, and the AIMS media and attribute lists.

The remaining day and a half was used for training and for conducting the test. The analysts were trained on the basic operation of AIMS, including hands-on development of a portion of the data base. During the second day, the analysts completed the AIMS data base by rating the media/attribute relations. AIMS work sheets for the DLIFLC training tasks/objectives were then filled out and entered into the instructional media selection routine. At the end of this session the analysts completed the AIMS portion of the user attitude questionnaire.

Training on ASTAR began the afternoon of the second day. It included creation of a subset of the actual database to be used in the test. Once the subjects became familiar with the procedures of ASTAR, they conducted an ASTAR Level 1 evaluation. The ASTAR Level 2 analysis was conducted on the morning of the final day. After finishing both ASTAR evaluations, the subjects completed the ASTAR portion of the user attitude questionnaire.

Results

ASTAR Evaluation. Two training systems, CMI and Programmed Text, were used for the ASTAR evaluations. The analysts performed an ASTAR Level 1 evaluation, as well as two ASTAR Level 2 analyses (Task and Subtask). The analyses compared each of the two training media to the operational system of foreign language reading comprehension for an intelligence analyst. The subjects input consensus ratings for each of the appropriate ASTAR questions for Level 1 and Level 2.

All three ASTAR evaluations indicated that CMI was predicted to be more effective than Programmed Text at training Spanish reading comprehension. The ASTAR Level 1 scores were 77.81 for CMI and 160.18 for Programmed Text. [Lower ASTAR summary scores indicate higher predicted training effectiveness.] For the ASTAR Level 2 analyses, the task summary scores were 81.95 for CMI and 102.03 for Programmed Text. The subtasks summary scores were 78.40 for CMI and 89.39 for Programmed Text. The differences in the summary scores between CMI and Programmed Text became smaller as the ASTAR level of evaluation increased.

The transfer problem appeared to be the major cause for the disadvantage with Programmed Text. This subscore was 40.00 for the ASTAR Level 1 analysis, which was over three times higher than the predicted transfer problem for CMI. The predicted transfer problem was lowered when more specificity was introduced with ASTAR Level 2.

AIMS Evaluation. The AIMS analysis was conducted with a pool of 17 media and 46 attributes. Of the potential pool of 17 media, a subset of 14 media were selected by AIMS as acceptable media for the 10 reading comprehension learning objectives. Of the 14 media, 7 were selected as possible media for all 10 of the learning objectives. In addition, 3 were selected for 5 of the

10 learning objectives, two were selected for 3 of the learning objectives, 1 was selected for 2 learning objectives, and 1 was selected for only 1 learning objective. The AIMS routine selected between seven and fourteen media for each objective. CMI had the highest rating for every objective. Three media were not selected as acceptable media: checklist, audio tape, and model.

Reactions To ASTAR. The analysts' overall reactions to ASTAR were that ASTAR was difficult, frustrating, and rigid. The inability to modify rating scales was the cause for the analysts' feelings that ASTAR was rigid. The analysis results were also found to be slightly difficult to understand because of the inability to interpret the numbers. However, the analysts commented that the language used in the program was consistent and understandable. Both subjects agreed that with an updating of the data entry procedures, ASTAR could be useful and more user friendly.

The analysts both felt that they could do work with ASTAR, but not very comfortably. Overall, they did not expect to have much use for ASTAR in their daily work of curriculum development. They also did not feel ASTAR could help with their job effectiveness.

Perceived positive aspects of ASTAR included:

ASTAR screens used clear labels in menus and the questions were clearly presented.

The terminology in ASTAR was consistent and easily understood.

The ASTAR instructional materials were very thorough.

Memory requirements during the use of ASTAR were very low.

Perceived negative aspects of ASTAR included:

Learning to use ASTAR was difficult.

During the use of ASTAR, tasks couldn't be performed in a straightforward manner, and feedback/error messages were not helpful.

ASTAR outputs were not useful. They were difficult to understand and the format was confusing.

Other items, including the organization of menus, helpfulness of prompts, ease of learning, and helpfulness of the instructional materials, were perceived as average.

Reactions To AIMS. Overall, the analysts seemed very pleased with AIMS. They found AIMS useful, easy, satisfying,

flexible, quite productive, and adequately powerful. The analysts stated that they could work comfortably with AIMS, and they believed that it would increase their job effectiveness. However, one of the analysts stated that he did not expect to have much use for AIMS in his daily work, and that whatever could be done with AIMS, he could do some other way. The only aspect of AIMS rated below average was the helpfulness of the AIMS error messages. It was given a rating of 3.5 on a scale with 1 = low and 7 = high. Almost all aspects of AIMS were rated well above average.

Conclusions

This operational test of ASTAR and AIMS provided a unique application of the two techniques. It permitted the analysts to determine whether the two DETs have sufficient merit to be used in the curriculum development processes. It also illustrated that the two DETs have applicability to a wide variety of training applications.

Both ASTAR and AIMS are relatively old programs. They were developed before the recent advancements in software design/technology and human computer interface design. The overall acceptance of ASTAR was rather low. On the other hand, AIMS was thought of quite highly. The user interface was the major factor in determining user acceptance. AIMS was generally acceptable, but certain features need to be improved before widespread adoption could be expected. The data base structure, data entry, and data editing procedures need revision within AIMS. ASTAR requires more extensive enhancements. ASTAR's biggest problems were the current rigid, unfriendly data base development and editing procedures. Simultaneous evaluation of devices was also a major concern for the users because of the time involved. Finally, ASTAR needs better data output options to make it easier for the analyst to visualize the data.

LONGITUDINAL TEST

This test of ASTAR was conducted using the Marine Corps M60A1 main battle tank as the weapon system of interest. It was an extension of Operational Test #1. The purpose of this test was to evaluate the utility and user attitudes of the ASTAR technique when exercised over an extended period of time. The test applied ASTAR to the comparative evaluation of the MCTFIST, GUARD FIST I, and SIMNET as potential training devices for the M60A1 tank. The test spanned a total of seven months, during which ASTAR was used to evaluate three training devices.

When this test was initiated, the developers of the MCTFIST were responding to a query of whether an alternative training device, such as GUARD FIST I, SIMNET, or U-COFT, could be used to adequately meet the MCTFIST training objectives. This longitudinal test was designed to compare the training effectiveness of MCTFIST, GUARD FIST I, and SIMNET on a common subset of the Marine M60A1 main battle tank training objectives.

The longitudinal aspect of the test follows a case history approach. User attitude questionnaires were obtained from subjects as they repeatedly applied ASTAR to the differing training devices. Self initiated logs, actual results derived from the use of ASTAR, and a termination interview were used to gather data. Descriptions of MCTFIST and GUARD FIST I were provided earlier. The description of the third device, SIMNET, is provided below.

SIMNET

SIMNET is an advanced, high technology, research and development program designed to explore a brand new technology consisting of large scale interactive SIMulator NETworking (SIMNET). The war fighting system undergoing test by the Army is a test bed to evaluate the ability of these technologies to support large-scale land battle collective (force-on-force) maneuver training.

When staffed appropriately, SIMNET allows force-on-force engagements from platoon to battalion task force level and provides training of selected command, control, combat support, and combat service support tasks at battalion level. Each SIMNET training device is created as a live interactive vehicle within a common simulated training environment. The SIMNET device used for evaluation in this test was configured to simulate the M1A1 main battle tank.

Method

Subjects. Two subjects participated in all phases of the test: a Project Director from NAVTRASYSCEN familiar with training analysis and design, and a contractor representative from the simulator manufacturer familiar with the device and the tank. These subjects were supplemented in the MCTFIST application by a tank gunnery sergeant who served as a Subject Matter Expert (SME). In the GUARD FIST I evaluation, a Project Director from PM TRADE assisted in the analysis. With the exception of the MCTFIST SME, all subjects were highly experienced in the use of computers.

Procedure. ASTAR evaluations of the three candidate training devices were conducted independently over a period of seven months. All trainers were subjected to ASTAR Level 1 and Level 2 evaluations. Two separate MCTFIST applications of ASTAR were conducted. During the course of the first MCTFIST test, subjects raised the concern that ASTAR was not designed for the multi-person crew trainer being evaluated. The investigators developed a modified version of ASTAR in which the questions and procedures were directed at multi-person crews. The changes involved assigning a equal portion of each appropriate question to each crew member, plus a communication factor. For example, on a 100 point scale 20 points were assigned to each of the four crew positions and 20 points were assigned to communication.

Following each ASTAR analysis, debriefing sessions were held with the subjects to discuss their experiences and opinions, and to identify any problems that would warrant future actions. In addition, the subjects were asked to complete an attitude questionnaire. After completion of the seven-month series of evaluations, a final termination interview was conducted with the Project Director from NAVTRASYS-CEN to assess his current views of ASTAR.

Results

Two MCTFIST Applications. The first MCTFIST analysis was conducted using the standard ASTAR program. The second analysis was conducted with the version of ASTAR modified for multi-person crew trainers described earlier.

The overall ASTAR summary scores for the two ASTAR Level 2 evaluations of MCTFIST showed virtually no difference; that is, they showed less than one point of variation on the ASTAR summary score. Subscore ratings showed a small difference between the two applications. The subscore rating for the "acquisition problem" was about 2.5 points higher with the multi-person crew version of ASTAR. However, the rating for "transfer problem" was approximately 1.8 points lower for the multi-person crew version of ASTAR. Hence, the net effect of the modified ASTAR was negligible.

It was not surprising to observe differences in the ASTAR subscores. The increased specificity from addressing the training objectives for each crew position could cause subjects to rate the acquisition problem as more complex. On the other hand, the decrease in the transfer problem may reflect a better concept of which tasks, for each crew member, would transfer adequately after completing training. Given the offsetting effects of this comparison, the need for an improved version of ASTAR to accommodate multi-person crew trainers seems debatable.

Comparison of the Three Devices. Both the ASTAR Level 1 and ASTAR Level 2 summary analyses predicted MCTFIST to be the most effective at training the Marine tank task requirements identified for this test. MCTFIST was followed by SIMNET and GUARD FIST I in order of predicted training effectiveness. The ASTAR Level 1 total scores were 66.70 for MCTFIST, 105.46 for GUARD FIST I, and 82.08 for SIMNET. [Lower ASTAR scores indicate higher predicted training effectiveness.] For the ASTAR Level 2 analysis, the total scores were 56.88 for MCTFIST, 96.19 for GUARD FIST I, and 77.70 for SIMNET.

The comparison of the summary scores between ASTAR Level 1 and Level 2 showed that all three trainers received lower scores at the Level 2 analysis. The differences in scores were not great, but they did reveal a general trend. This trend indicated that with increased supporting evaluation data and increased knowledge about the training device, a more accurate analysis of

effectiveness was reached.

The prediction that SIMNET would be more effective than GUARD FIST I seemed questionable, because GUARD FIST I was physically and functionally closer to MCTFIST. Subjects appeared to have trouble evaluating SIMNET because it was designed for a much different training objective. Therefore, examination of the acquisition and transfer subscores appeared justified. This examination revealed that the summary scores were driven by the transfer portion of the ASTAR score. Much of the problem subjects had in evaluating SIMNET seemed to focus on the transfer of tasks to the operational environment. Hence, this portion of the score was probably not valid. The acquisition portion of the ASTAR score reflected the expected ranking of the three devices. On the ASTAR Level 1 analysis acquisition subscore, MCTFIST was rated best with a subscore of 39.29, followed by GUARD FIST I at 46.69 and SIMNET at 55.0. The ASTAR Level 2 analyses showed the same ranking on the acquisition subscore.

User Attitude Questionnaire. The subjects were asked to complete the user attitude questionnaires following each application of ASTAR during the longitudinal test. Changes in perception across the series of applications were the point of interest. The comment below present the results of questionnaire items which reflected a change in user perception.

- ASTAR was initially perceived as difficult to use. After extensive use, subjects felt ASTAR was only moderately difficult to use.
- Users rated ASTAR to be less effective with continued use.
- Initial reactions to ASTAR rated the program as moderately useful, satisfying, and adequately powerful. By the end of the longitudinal test subjects expressed the opinion that ASTAR was rigid, frustrating and generally unproductive.
- With continued use, the organization of the ASTAR menus was viewed as more illogical and confusing.
- The consistency of the terminology and language used in ASTAR was rated better over time.
- Learning to operate and work with ASTAR was perceived to be more difficult with extended use.
- The ASTAR instructional materials were considered to be less helpful but more complete than on first inspection.

The aspect of ASTAR most often criticised was the format of the ASTAR results as presented in the final summary. It was felt that the lack of definition of the data rendered the summary

data meaningless. The general tenor of the comments indicated that the subjects did not know what the data were telling them. No documentation or screen presentations told them how to interpret the different scores.

The aspect of ASTAR which users liked most was the overall concept of an automated system to perform training effectiveness evaluations of multiple training devices. It was felt that the use of such a technique could result in cost, time, and manpower savings. They also liked having a tool to provide quantitative data which could be used in making decisions during the design and development process of training systems.

Termination Interview. An interview was conducted with the Project Director from NAVTRASYSCEN at the completion of the longitudinal test. The following section addresses his feelings about ASTAR.

The ASTAR system, in itself, did not help to increase user confidence. The continuing discussions and interactions with the experimenters and ASTAR SMEs increased the user confidence level. ASTAR user manuals were not used. They were too long and did not have a good summary format. A strong desire to better understand the end output was expressed. A method to help interpret the output data was needed. The ASTAR questions themselves were written and presented in a straight forward manner, but it was difficult to identify the items which were the most important in driving final output data.

Through continued use of ASTAR, the level of understanding and acceptance had definitely increased. The subject did not become discouraged with ASTAR over the entire seven months of the longitudinal test. Inherent benefits from using ASTAR were quite apparent right from the start. Even with its inherent problems, the subject still felt that ASTAR could easily become a valuable technique. If ASTAR were modified to improve the user interface, the subject would use it and reconsider the areas of feasible application. Further developmental changes to improve the interaction with the user would be highly supported.

The subject felt confident in teaching or explaining to a new user how to effectively use ASTAR. He became comfortable with applying ASTAR on his own in a new area. The subject also became familiar with the type of assumptions that need to be made to run the program. The power of ASTAR may lie in the forced and structured interaction of SMEs in making ratings and tradeoff decisions. It was suggested that if ASTAR was used in its present state, then a third party ASTAR expert should be present to help new analysts through the evaluation. The third party might help organize the data inputs.

Conclusions

This test provided an ongoing use and evaluation of ASTAR

over a seven month period. The M60A1 weapon system selected for the test, as well as the MCTFIST, GUARD FIST I, and SIMNET training devices, provided a good application of the technique. They permitted the analysts to compare the merits of the three simulators as M60A1 full crew tank trainers, as well as to evaluate the utility and usability of the decision aid.

Overall, the user acceptance of ASTAR was rather low. It was evident that ASTAR was considered quite worthwhile in concept. However, ASTAR was unsatisfactory in terms of user friendliness. The user interface was clearly the major factor in determining future acceptance of the two methodologies.

ANALYSIS OF OPERATIONAL TESTS

The subjects involved in the operational tests of ASTAR all felt that automated decision aids to assist instructional developers should be valuable tools. However, neither decision aid evaluated in these tests was acceptable in its present format.

The opinions of the two DETs were somewhat variable. ASTAR received generally negative ratings overall. However, the detailed items that were cited as problems varied from test to test. On the other hand, AIMS received generally positive ratings overall. However, just like ASTAR, there was considerable variability in what analysts liked and disliked about AIMS. Items that some analysts rated very highly were rated as very low by other analysts. The variability appears to be linked to the background and expertise of the various analysts across the spectrum of tests. Another contributing factor may have been the analyst's initial perceived need for each DET. Analysts who agreed to participate in the tests usually had expressed a higher need for one or the other of the two DETs in their initial survey response. They agreed to use both techniques during the tests, but they may have entered the test with some biases. If they preferred a technique, it was generally the one for which they had initially expressed a greater need. One clear conclusion emerged from the tests. The analysts' opinions and acceptance of each DET was driven by the perceived user friendliness and ease of use/learning. If the DET does not have a good user interface, then it will not be acceptable to users in the field.

Both DETs were cited as having desirable characteristics. For evaluating and comparing the training effectiveness of different emerging devices, or of proposed changes to existing devices, the availability of a DET such as ASTAR was seen as a valuable tool. For example, the ability to conduct ASTAR evaluations at three different levels of device development was felt to be of particular benefit. The analysts believed that the proper application of ASTAR should result in considerable savings in time and cost during the analysis phases of training development. Regardless of their negative opinion of ASTAR, the analysts clearly thought the concept was worthwhile, and that

there was a use for a DET like ASTAR.

ASTAR and AIMS are relatively old programs. They were developed before many of the recent advancements in the design and technology of both software and human/computer interfaces. AIMS was considered generally acceptable, but certain features still need to be improved to ensure widespread adoption. The primary needs in AIMS are an improved data base structure, an improved data entry capability, and an improved editing function. ASTAR will require much more extensive enhancement in the same general areas as AIMS before it can gain general user acceptance. ASTAR's greatest needs are for a better data base development capability and a more systematic, expanded editing function. ASTAR also needs better data output options (graphics perhaps), and a system of "helps" or explanations of the summary data, to make it easier for the analyst to interpret test results.

The findings of these operational tests indicate that both ASTAR and AIMS will require modifications before implementation as standard evaluation techniques. Without these modifications, user acceptance will be poor at best. Most of the shortcomings can be alleviated by modifying the programs to incorporate current software practices, data base techniques, and user interface standards. The subjects who participated in the evaluation made a number of specific recommendations for improvement to both the ASTAR and AIMS interfaces. Given the nature of the required modifications, an acceptable version of the programs should be achievable with only moderate resources. However, there would still be areas of concern that would be more difficult to alleviate. Terminology, for instance, tends to be application-specific, so it would be difficult to use generic terminology in the ASTAR questions and prompts. In addition, changes to the basic ASTAR computational formulas could not be made without negating previously established validity. As a result, shortcomings associated with the mechanics of ASTAR, as identified by the analysts who participated in these operational tests, could be corrected, while content-related flaws could not be easily addressed.

One additional problem associated with ASTAR was finally identified during Operational Test #3. Over the three operational tests, it was observed that problems would occasionally be experienced in attempting to complete an ASTAR Level 2 or 3 analysis. Access to the data base during the use of ASTAR would occasionally cause the program to "hang up" or not accept data. Based on the computers used in this test, it became apparent that the compiler used for ASTAR is outdated. Looking across the run problems in the three tests, it was observed that ASTAR always ran correctly on an 8088 or 8086 based computer. However, when ASTAR was run on a 80286, AT class of computer, data base problems were always encountered. It appears that the compiler used for ASTAR does not generate code that is totally 80286 compatible. This is unacceptable for any implementation plan and requires ASTAR to be recoded to avoid this problem in the future.

The most outstanding feature found in ASTAR was the structured group approach to training device design that it suggests. Ideally, ASTAR is conducted by multiple raters from differing fields, who are familiar with the training design process. The suggested mix of personnel includes three to five persons from these disciplines: Instructional Technology, Psychology, Engineering, Human Factors, and Subject Matter Expertise. ASTAR asks each member of the design team to review the training device design individually and then to meet to discuss the major assumptions regarding the eight major ASTAR analyses. The raters conduct their analysis independently and record their reasoning behind the chosen ratings. The team then meets again to discuss and to compare results and to determine the reasons behind any difference in judgements. The raters reassess their judgements, striving for consensus. This iterative meeting process becomes an exercise in compromise until all members find an acceptable design.

Using ASTAR in this method involves each member of the design team and ensures that suggestions from Human Factors, Instructional Technologist, and Design Engineers are considered throughout the design process. This approach helps clarify the assumptions made by different disciplines. Often design team members have difficulty in communicating the importance of contributions available from their individual disciplines. ASTAR facilitates communication between the diverse disciplines responsible for training device design by providing the necessary platform.

Summary of User Comments

User comments on ASTAR ascertained in the user attitude questionnaire provided the following insights.

1. The most-liked aspects of ASTAR:

- a. The overall concept of an automated system to perform training effectiveness evaluations of multiple training devices. This concept was considered quite worthwhile. It was felt that the use of such a technique could result in cost, time, and manpower savings.
- b. Having a tool to provide quantitative data which can be used in the decision making process during the design and development of training systems.
- c. Computer documentation of trainer and weapon system hardware, controls and displays, and operator tasks on IBM compatible software.

2. The least-liked aspects of ASTAR:

- a. The output data, or ASTAR results, as presented in the final summary. It was felt that the lack of definition

of the data rendered them meaningless. The general tenor of the comments indicated that the subjects did not know what the data were telling them, and there were no documents or screen presentations to tell them how to interpret the different scores.

- b. The tediousness and length of time associated with the entry of almost identical lists of controls and displays. This requires entries for the operational system and the trainer in both the workbook and the computer. This was believed to be unnecessary, redundant, and inefficient.
 - c. The lack of organization of the menus which prohibited a free flow in and out of the process. In other words, there was no capability to escape from the program at any point and then return at a later time to the same point. This could be done, of course, but not quickly and conveniently. Instead, the user was forced to work through a time consuming and complex procedure to arrive at a point of interest.
3. User-suggested improvements to ASTAR:
- a. Reprogramming ASTAR to make it more user friendly and to provide a more meaningful output, e.g. graphics outputs.
 - b. Making the system more user friendly by adding the following capabilities:
 - (1) Providing simplified utility menus to allow easy editing, addition and deletion of controls/displays, and task and subtask data.
 - (2) Providing a way to save data on both hard drive and floppy disks.
 - (3) Providing input/output capabilities from database and spreadsheet programs;
 - (4) Allowing revision of data base;
 - (5) Allowing input to be duplicated;
 - (6) Upgrading to mouse input; and
 - (7) Allowing side-by-side comparison of two systems rather than the current practice of producing output for one system followed by output for the next.

PAG Assessment

The PAG met after the operational tests were completed. The purpose of the PAG meeting was to assess the findings of the evaluation. The PAG reviewed:

- a summary of the evaluation tests;
- the PAG Assessment Objectives of ASTAR impact, cost,

and development time; and

- the recommended Functional Description.

The conclusions of the PAG were mixed. Although the concept and underlying benefit of ASTAR was recognized, the current state of the software overshadowed any benefit to be derived by recommending that it be distributed or institutionalized. Users' opinions indicate they do not value ASTAR in its present software configuration. ASTAR was perceived to be inadequate as it stands. The comments of the PAG were unanimous in suggesting a new start to incorporate the "concept", "approach", or "philosophy" of ASTAR into an improved software package.

TRANSITION/IMPLEMENTATION PLAN

This transition/implementation plan was developed to guide the transfer of ASTAR to DoD agencies involved in the acquisition of training devices. It addresses both the current version of ASTAR and an improved ASTAR labeled ASTAR II. ASTAR II was designed to alleviate the users' concerns found during the course of this operational evaluation. The requirements for ASTAR II are described in Volume III: Functional Description. Methods for using ASTAR in conjunction with other DETs is discussed in the Analytic Investigation section of Volume II: Test Reports.

NEED FOR DETS

Training device designers have available to them many hardware options to satisfy the training requirements of the operational environment. The designer's task is to make tradeoff decisions between technology and instructional features to improve training effectiveness at the lowest cost. The determination of a device's potential training effectiveness, and of its associated costs, need to be made early in the acquisition process to optimize the design of the total training system (Martin, Rose, & Wheaton, 1988).

DETs are designed to aid the training device designer in determining optimum training device configurations. Methods are needed to assist training device designers by standardizing the tradeoff process. The utilization of standardized DETs can identify the design features, given cost constraints, which lead to the greatest amount of transfer of training and restrict the number of configuration options required to be examined. The goal of DETs is to relieve the designer of tedious tasks and to allow the effort to focus on the important issues of determining the design specification. Optimal device configuration options may be quickly determined if effective DETs are employed.

To empirically determine if a need exists within the user community, a survey was conducted (Bradley, 1990). The survey was distributed through government organizations and conferences to 183 potential users, with 46% of the surveys returned. The respondents were asked to indicate whether they had a definite need, a possible need, or no need at all for the analytic DETs of ASTAR and AIMS. Overall, there was a definite dichotomy of responses. The majority of responses expressed either high interest in DETs or no interest at all. In general, if respondent were interested in DETs, then they were interested in both ASTAR and AIMS. The returned surveys indicated that 52.5% and 59.0% of the respondents had a positive interest in ASTAR and AIMS, respectively. This majority interest in ASTAR and AIMS indicated a strong desire and perceived need for design aids by operational analysts.

ASTAR SUPPORT OF THE INTERSERVICE PROCEDURE FOR INSTRUCTIONAL SYSTEMS DEVELOPMENT (IPISD)

An evaluation conducted by Martin, Rose, and Wheaton in 1988 outlined the training device acquisition process of the Armed Services to suggest ways in which ASTAR could be used in these processes to facilitate the acquisition of effective devices. They suggested points in each service's training device acquisition process (Army, Navy, and Air Force) where ASTAR could be utilized. In an effort to continue the analysis of how ASTAR could be used to improve the effectiveness of the device acquisition cycle, this portion of the implementation plan will illustrate the usefulness of combining the ASTAR evaluation technique with the Interservice Procedure for Instructional Systems Development.

ASTAR may be useful in three of five IPISD phases. Phase I of the IPISD process, Analyze, indicates ASTAR will work effectively in four out of the five blocks of the Analyze phase: Select Task/Functions, Construct Performance Measures, Analyze Existing Courses, and Select Instructional Settings. Phase II of the IPISD process, Design, provides much of the qualitative data needed for conducting ratings in the three ASTAR analysis levels. Two of the four blocks within Phase II, specifically Develop Objectives and Describe Entry Behavior, provide information to ASTAR. In Phase II, Develop, ASTAR can be used during the Review Existing Materials and Develop Instruction blocks. Within Phase III, ASTAR makes a direct contribution to the ISD process. Phases IV and V of the IPISD process, Implement and Control, do not relate directly to known uses of the ASTAR technique.

The following summarizes some of the major areas within the IPISD process where ASTAR could be used to assist the training device designer, and where IPISD outputs could be utilized as data inputs by ASTAR:

I. ASTAR uses within IPISD:

- a. Examine training effectiveness of existing materials.
- b. Structure development of training objectives.
- c. Document procedures and major decisions derived.
 1. Document the rationale used in the exclusion and inclusion of the media alternatives.
 2. Document the rationale on a task-by-task basis by which existing courses are excluded from consideration.
- d. Support development of device scripts.
 1. Develop task level device configurations for scripts.
 2. Develop specific control and display configurations.
- e. Iteratively examine alternative tradeoff solutions.

II. Data input provided by IPISD:

- a. Edited task lists.
- b. Performance level expected from training.
- c. Knowledge levels and skills necessary.
- d. Entry characteristics of the trainee.
- e. Material, procedures, plans, and media necessary to conduct instruction.

IMPLEMENTATION OF ASTAR

An optional approach to the implementation of ASTAR is to implement the concept of ASTAR. Under this option, the implementation directive would require that the variable categories, i.e., performance deficit, learning difficulty, etc., be considered in the design of training systems and estimations of device effectiveness. ASTAR could be used as one method, but not the only method, of satisfying the requirement. This approach may be desirable even if an improved version of ASTAR is eventually developed.

Limitations

It may be necessary to implement ASTAR in a limited domain to users with a critical need. It should be noted that the current version of ASTAR, as it is presently compiled, is not fully compatible with PC/AT or later classes of MS-DOS computers. Errors may be encountered when attempting to conduct a complete ASTAR Level 2 or Level 3 analysis. ASTAR is capable of running without error on a PC or PC/XT only.

Configuration Control

ASTAR software has been distributed by this evaluation and will undoubtedly be redistributed within the DoD and Military Training Community. Configuration Control, therefore, must be initiated by assigning a responsible Agency within DoD the managerial responsibility for distribution control of the ASTAR software and user's manuals. Any modifications made in the basic formulas within the ASTAR data base should be made only by that designated agency. The internal design of ASTAR should be such that modifications to the basic formula cannot be made except by the assigned configuration control manager.

Transition Materials

Transition materials will be addressed for both ASTAR and ASTAR II. Though ASTAR is not recommended for general implementation, transition materials were provided for ASTAR. Some individuals may have sufficient need for ASTAR to use it regardless of its current shortcomings.

ASTAR. Transition materials for the current ASTAR program are a 62 page ASTAR User's Manual prepared by the American Institutes for Research and a 20 page ASTAR Abbreviated User's

Manual developed by UCF/IST. A series of Briefing/Training Slides were prepared to assist new or potential users in the application of ASTAR during this evaluation. All of these materials, and the current ASTAR software, are available from the NAVTRASYSCEN, Research and Engineering Department.

ASTAR II. Transition materials for projected users of ASTAR II would be designed to include off-line/on-line manuals and on-line tutorials. The contents of disk-based "read me" file could be printed to provide supplemental hard copy material. An improved set of briefing slides reflecting the features of ASTAR II would be developed.

Training Approaches

ASTAR II Training. ASTAR II large group training materials could be created from the on-line tutorial to provide an improved set of Briefing/Training Slides applicable to ASTAR II. Primary training would be conducted on an individual basis through an on-line tutorial contained within ASTAR II. It is envisioned that the improved design identified for ASTAR II would greatly reduce the level of required training.

RECOMMENDATIONS

This ASTAR evaluation was sufficiently comprehensive to make a thorough assessment of the projected user acceptance of the current or projected update to the ASTAR program. A combination of user comments, test observations, and PAG determinations were used to evolve these recommendations concerning ASTAR. Recommendations will be presented for the current ASTAR program, recompiled ASTAR, and ASTAR as it could exist if improved according to the Functional Description for ASTAR II contained in Volume III: Functional Description.

CURRENT ASTAR

The current ASTAR was found to be so user unfriendly as to make its use counterproductive. ASTAR, as it exists, is not recommended for use as a standard within DoD activities. Its use as a stopgap means of comparing training systems should be undertaken only with full knowledge of the difficulties to be met.

RECOMPILED ASTAR

The current version of ASTAR could be recompiled on a newer version of a COBOL compiler to upgrade ASTAR for limited interim use, pending a redesign of ASTAR. This recompiling of ASTAR would permit ASTAR to be used on all currently available MS-DOS microcomputer, but it would still have all of the user problems encountered in the current ASTAR.

ASTAR II

It is recommended that any further developmental action on ASTAR be limited to consideration of either the design described in the Functional Description contained in Volume III: Functional Description or a totally new effort. The functional description for ASTAR II is the minimal response to the problem areas identified in this evaluation. It would provide a design approach which would satisfy user demands for a modern aid to device development.

UNRESOLVED ISSUES

The ASTAR II addresses the user interface issues found during the course of this evaluation. However, a number of fundamental content issues remain which were not addressed in ASTAR II. Any attempt to address content issues will require a major effort to develop a system for estimating training effectiveness.

Major remaining issues are:

1. Many users would like to use ASTAR to compute an absolute index of goodness. ASTAR was not designed for this. Therefore, this feature cannot be accommodated

without a complete rework of ASTAR.

2. The basic formulas for ASTAR are still questioned. The original predictive validity was low, and people do not understand how they were derived.
3. ASTAR considers hardware to be the critical factor. ASTAR does not make any provisions for considering such training materials as scenarios, which seems to be a major shortcoming in today's training environment.
4. ASTAR is basically oriented to single person training devices. What about crew trainers? These are becoming increasingly important. The current ASTAR requires crews to be handled through an increase in the task and subtask data base. But ASTAR requires a restricted task set to make it manageable, especially at Level 3. This creates an obvious conflict.

AIR RECOMMENDED CHANGES

The idea that ASTAR needs improvement is not new. As part of the final report on DEFT and ASTAR (Rose & Martin, 1988), AIR developed a set of recommended modifications to ASTAR. Their recommendations correspond closely to those which have resulted from this research project and which are addressed in the design of ASTAR II. The AIR recommendations and discussions are replicated below as supporting information.

ASTAR could be improved in several ways. We have generated suggested improvements to ASTAR throughout our series of evaluations. These improvements will be detailed below. They include improving the wording of the scales, making the input procedures more user friendly, publishing a user's manual, and standardizing rating procedures.

We have generated some suggested modifications to DEFT/ASTAR based on our past experiences and discussions with potential users of the program. These are described below; we will avoid confusion by referring to the program by its current name, ASTAR.

1. Resolve inconsistencies in the three levels of analysis; there are some scales where the ratings are qualitatively different. For example, Performance Deficit analysis asks the analyst the proportion of skills/knowledge the trainee must learn in ASTAR 1 and ASTAR 2. ASTAR 3 asks the analyst for a 0 to 4 rating (essentially go or no-go) for each subtask.
2. Change the (1) Performance Deficit, (2) Learning Difficulty, (4) Residual Deficit, and (5) Residual Learning Difficulty analyses to ratings of each

skill and knowledge required to perform the training and operational tasks. For ASTAR 1, the analyst would rate each of the skills and knowledge requirements and the program would combine these ratings to produce a summary score. For ASTAR 2 and 3, the program would combine the analyst's skill/knowledge ratings to produce scores for each task/subtask and a summary score. Similarly, change the (6) Physical and (7) Functional Similarity scales so that controls and displays are rated once and then combined for tasks, subtasks, and a summary score.

3. For ASTAR 3, decide where a subtask level of analysis is too detailed and change the program so that ratings are done on a task level. We are certain this needs to be done for the (3) Acquisition Efficiency and (8) Transfer Efficiency analyses, and it may be appropriate for other scales as well (Performances and Residual Deficit, Difficulty).
4. Refine the scale definitions and provide more and better anchor points. Decide whether 10-point scales, (or some other type) are more appropriate than 100-point scales and change the program if necessary. Bound the summary numbers (possibly using a maximum of 100), and reverse the direction so that "better" device score is a higher number rather than a lower number.
5. Improve methods for developing databases and entering ratings.
6. Write the program in a language other than COBOL so that it will run faster.
7. Provide for various versions of the summary screens. For example, provide a summary of the analyst's ratings for each of the eight analyses on the summary screens of ASTAR 2 and 3. Also, improve the presentation of the listing of individual ratings, and make it easier for the analyst to change ratings.
8. Create beginning and advanced user versions of the program. We envision the beginner's version as having very detailed explanations of the scale on the screen and the more advanced version as similar to a spread sheet format.
9. Add the capability to calculate average ratings for multiple judges.
10. Add a mechanism to record on-line analyst's

assumptions and reasons for ratings.

11. Rewrite the screens and replace the psychological terms with words more familiar to the lay person, or define those terms if that is more appropriate. Restructure the screen and sentences to make them easier to read.
12. Revise the user's manual to incorporate the program changes and to make it easier to use. These revisions would include specifying standard procedures for describing tasks, skill/knowledge requirements, and controls/displays, and for determining ratings.

These proposed revisions fall into three general types. The first type is revisions that involve conceptual changes to the model. Specifically, numbers 1,2,3, and part of 4 involve fundamental modifications of the ASTAR logic and underlying algorithms. The second type is revisions of the software that do not alter the underlying logic. Numbers 5 through 10 fall into this group; all would require extensive software modifications. The third type includes editorial revisions (number 11) and external supplement (number 12).

Although the first four changes are reasonable, the implementation of them is beyond the scope of this project. Further developmental and validation research would be required before we would change the underlying logic. The current version of ASTAR still represents a significant advance over current methods of device effectiveness forecasting; we have no justification for changing the model or its conceptual underpinnings.

Implementation of the second type of changes -- involving extensive software revisions -- are also beyond the scope of this project. We will eventually make these changes, but not at the expense of this project.

The remaining changes -- rewriting the screens and producing a User's Manual -- Have been accomplished. They are presented in the following sections.

(Rose & Martin, 1988, pp.9-12)

BIBLIOGRAPHY

- Adams, A. V., & Rayhawk, M. (1987). Review of Models of Cost and Training Effectiveness Analysis (CTEA) Volume 2 (Report No. ARI-RN-87-59). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.
- Allbee, K. E., & Semple, C. (1981). Aircrew Training Devices: Life Cycle Cost and Worth of Ownership (Report No. AFHRL-TR-80-34). Brooks AFB, TX: Air Force Human Resources Laboratory.
- Allison, S. L. (1969). The Pilot Training Study: A Cost-Estimating Model For Undergraduate Pilot Training (Report No. RM-6083-PR). Santa Monica, CA: Rand Corporation.
- American Institutes for Research (1987). Review of the DEFT Technique and Recommendations for the Integration of DEFT into the Training Device Design Process. Washington, D.C.: American Institutes for Research.
- American Institutes for Research (1988). ASTAR User's Manual. Washington, D.C.: American Institutes for Research.
- Billings, J. (1983). User's Manual for Training Device Cost Model 'TRACOM' [Final report] (Report No. PMT-EM-0004-83). Orlando, FL: Science Applications, Inc.
- Bird, R.G. (1989). Automated Instructional Media Selection (AIMS) Users Manual. Orlando, FL: Engineering and Economics Research, Inc.
- Bird, R., Gibbons S., & Companion, M. A. (1990). Operational Test #1: M60A1 Main Battle Tank. Technical Interim Report. Orlando, FL: University of Central Florida/Institute for Simulation and Training.
- Blaiwes, A. (in press). AIMS. Proceedings of the NTSC/IST Cost Effectiveness Analysis Workshop. Held in conjunction with the 11th Interservice/Industry Training Systems Conference, Ft. Worth, TX.
- Boren, Jr, H. E. (1969). The Pilot Training Study: A User's Guide to the Advanced Pilot Training Computer Cost Model (APT) [Study report Vol. 8] (Report No. RM-6087-PR). Santa Monica, CA: Rand Corporation.
- Braby, R., Henry, J. M., Parrish, Jr., W. F., & Swope, W. M. (1975). A Technique for Choosing Cost-Effective Instructional Delivery Systems (Report No. TAEG-16). Orlando, FL: Naval Training Equipment Center.
- Braby, R., Micheli, G. S., Morris, Jr., C. L., & Okraski, H. C. (1972). Staff Study on Cost and Training Effectiveness of Proposed Training Systems (Report No. TAEG-1). Orlando, FL:

Naval Training Equipment Center.

- Bradley, B. (1990). Survey Results. Technical Interim Report. Orlando, FL: University of Central Florida/Institute for Simulation and Training.
- Bradley, B., & Companion, M. (1989). ASTAR: The Automated Simulator Test and Assessment Routine. Proceedings of the Cost Effectiveness Analysis Workshop. Held in Conjunction with the 11th Interservice/Industry Training Systems Conference.
- Bradley, B. A., & Companion, M. A. (1989). The Automated Simulator Test and Assessment Routine (ASTAR) Abbreviated User's Manual. Orlando, FL: University of Central Florida/Institute for Simulation and Training.
- Bradley, B., & Companion, M. (1990). Operational Test #3: Foreign Language Reading Comprehension. Technical Interim Report. Orlando, FL: University of Central Florida/Institute for Simulation and Training.
- Calabro, J. R., & Pozzo, M. M. (1983). CEDA (Capability for Evaluating Decision Aids) High-Level Software System Design [Technical report] (Report No. MTR-8651; ESD-TR-82-419). Hanscom AFB, MA: Electronic Systems Division.
- Calabro, J. R., & Pozzo, M. M. (1983). CEDA (Capability for Evaluating Decision Aids) User's Manual [Technical report] (Report No. MTR-8658; ESD-TR-82-418). Hanscom AFB, MA: Electronic Systems Division.
- Caver, T. V. (1977). Training Developments: A Means to Reduce Life Cycle Costs [Study project report]. Fort Belvoir, VA: Defense Systems Management Coll.
- Clark, D. A., Levine, J. G., Zunic, G. J., & Gardner, R. L. (1985). Evaluation of the C-E Cost Allocation Algorithms Volume 6 (Report No. TR-118-7). State College, PA: Desmatics, Inc.
- Companion, M. A. (1989). ASTAR: Operational Evaluation and Implementation Work Plan Report (Contract No. 61339-89-C-0029). Orlando, FL: University of Central Florida/Institute for Simulation and Training.
- Companion, M., & Bailey, M. (1990). Operational Test #2: SEAWOLF Internal Auxiliary Launcher. Technical Interim Report. Orlando, FL: University of Central Florida/Institute for Simulation and Training.
- Connelly, M. N. (1967). A Personnel Cost Data Bank for Use in Studies of Cost and Effectiveness [Research memo.] (Report No. NPRA-SRM-68-3). San Diego, CA: Naval Personnel Research Activity.

- Donnell, M. L., & Adelman, L. (1980). Development of a Computerized Training Requirements and Cost Evaluation System for the U.S. Marine Corps [Semi-annual report]. McLean, VA: Decisions and Designs, Inc.
- Donnell, M. L., Adelman, L., & Patterson, J. F. (1980). Development of a Computerized Training Requirements and Cost Evaluation System for the U.S. Marine Corps [Final technical report]. McLean, VA: Decisions and Designs, Inc.
- Dynamics Research Corporation (1988). Joint Service Instructional Systems Development/Logistics Support Analysis Record decision Support System (ISD/LSAR DSS). Detailed Functional Specification. Volume 1-Technical. Wilmington, MA.
- Gardner, P. R. (1984). Computer Based Training Cost-Benefit Model (Report No. HEDL-7401). Washington, DC: Department of Energy.
- Gibbons, S., & Companion, M. (1990). Longitudinal Test: M60A1 Main Battle Tank. Technical Interim Report. Orlando, FL: University of Central Florida/Institute for Simulation and Training.
- Gibson, R. S., & Orlansky, J. (1988). Macro-Level Training Data: A Conceptual Model Relating Training Cost and Effectiveness Data to Training Policy Interests and Issues (Report No. IDA-P-2042). Alexandria, VA: Institute for Defense Analyses.
- Goldberg, I. (1988). Training Effectiveness and Cost Iterative Technique (TECIT) Volume 1 (Report No. ARI-RN-88-35). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.
- Goldberg, I., & Khattri, N. (1987). Review of Models of Cost and Training Effectiveness Analysis (CTEA) (Report No. ARI-RN-87-58-VOL-1). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.
- Hartung, W. G. (1969). Use of Computerized Cost Models in Cost Analysis (Report No. ESD-TR-69-120). L G Hanscom Field, MA: Electronic Systems Division.
- Hays, B., & Singer, M. (1988). Simulation fidelity in training systems design; Bridging the gap between reality and training. New York: Springer-Verlag.
- Instructional Science and Development, Inc. (1984). Automated Instructional Design System (AIDS) user manual. San Diego, CA: Author.

- Interservice Procedures for Instructional Systems Development: Executive Summary and Model (1975) (TRADOC Pamphlet 350-30). Fort Benning, GA: U.S. Army Combat Arms Training Board.
- Joules, R., & Geissler, P. (1983). Military Occupational Speciality Training Cost Handbook (MOSB). Indianapolis, IN: Army Finance and Accounting Center.
- Kantowitz, B. H., & Roediger, H. L. (1984). Experimental psychology; Understanding psychological research (2nd ed.). New York: West Publishing Company.
- Kennedy, R. S., Jones, M. B., & Baltzley, D. R. (1988). Empirical Demonstration of Isoperformance Methodology Preparatory of an Interactive Expert Computerized Decision Aid (Report No. ARI-RN-88-93). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.
- Kildebeck, J. S., Kipnis, G. M., & Mackey, W. E. (1969). Rated Resource Requirements System. A Linear Programming Determination of a Least-Cost Pilot Structure for a Simulated Environment (rev. ed.). Washington, DC: Directorate of Manpower and Organization (Air Force).
- Klein, G. A., Johns, P., Perez, R., & Mirabella, A. (1985). Comparison-Based Prediction of Cost and Effectiveness of Training Devices: A Guidebook (Report No. ARI-RP-85-29). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.
- Knapp, M. I., & Orlansky, J. (1983). Cost Element Structure for Defense Training (Report No. IDA-P-1709; IDA/HQ-82-25216; SBI-AD-E500 628). Alexandria, VA: Institute for Defense Analyses.
- Knerr, B. W. (1984). Instructional Materials: Life Cycle Cost Analysis for Human Factors, Manpower Personnel and Training Issues in Military System Development (Report No. ARI-RN-84-140). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.
- Knollmeyer, L. E. (1969). The Pilot Training Study: A Cost Estimating Model for Advanced Pilot Training (APT) [Study report Vol. 7] (Report No. RM-6086-PR). Santa Monica, CA: Rand Corporation.
- Kribs, D. (1989). Automating Instructional System Development. San Diego, CA: Instructional Science and Development, Inc.
- Kribs, H. D., & Mark, L. J. (1982). Evaluation of Media Selection [Special report] (Report No. NPRDC-SR-82-13; SBI-AD-F630 628). San Diego, CA: Navy Personnel Research and Development Center.

- Liberati, G. L., Egber, D., & French, J. (1985). ASSET (Acquisition of Supportable Systems Evaluation Technology) Users Guide: Application (Report No. AFHRL-TP-85-25(1)). Brooks AFB, TX: Air Force Human Resources Laboratory.
- Littleton, L. (1969). The Pilot Training Study: A User's Guide to the Undergraduate Pilot Training Computer Cost Model (Report No. RM-6084-PR). Santa Monica, CA: Rand Corporation.
- Littleton, L. (1969). The Pilot Training Study: A User's Guide to the Pilot Computer Model [Study report Vol. 2]. Santa Monica, CA: Rand Corporation.
- Marcue, N. C., Blaiwes, A. S., & Bird, R. G. (1982). Computer Aided System for Developing Aircrew Training (CASDAT) an Automated Aid for Developing Aircrew Training (Report No. NAVTRAEQUIPCEN 79-C-0076-1). Orlando, FL: Naval Training Equipment Center.
- Mark, L. J., Kribs, D., Schuler, J. W., & Brown, J. E. (1982). A Life-Cycle Cost Model for an Electronic Maintenance Trainer (Report No. NPRDC-SR-82-33). San Diego, CA: Navy Personnel Research and Development Center.
- Martin, M. F. (1987). Implementation of ASTAR: Evaluation Plan for the Combat Talon II Maintenance Trainer. Washington, D.C.: American Institutes for Research.
- Martin, M. F., & Rose, A. M. (1988). Implementation of ASTAR: Evaluation of the Portable Aircrew Trainer. Washington, DC: American Institutes for Research.
- Martin, M. F., Rose, A. M., & Wheaton, G. R. (1988). Applications for ASTAR in Training System Acquisition. Washington, DC: American Institutes for Research.
- Matlick, R. K., Berger, D. C., Knerr, C. M., & Chiorini, J. R. (1980). Cost and Training Effectiveness Analysis in the Army Life Cycle Systems Management Model (Report No. ARI-TR-503). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.
- Matlick, R. K., Berger, D. C., & Rosen, M. H. (1980). Cost and Training Effectiveness Analysis Performance Guide (Report No. ARI-RP-81-1). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.
- Matlick, R. K., Berger, D. C., & Rosen, M. H. (1980). Cost and Training Effectiveness Analysis (CTEA) Performance Guide [Final report]. Arlington, VA: Litton Mellonics Systems Development Division.
- May, L. J., & Mayberry, P. W. (1986). Rand Cost-Performance Model for Setting Qualification Standards: Preliminary

Comments (Final report) (Report No. CRM-86-200). Alexandria, VA: Center for Naval Analyses, Marine Corps Operations Analysis Group.

McCannel, G. J. (1985). Cost Determination Model for the Functional Context Training Revision of Basic Electricity and Electronics Training. Master's thesis, Naval Postgraduate School, Monterey, CA.

Morris, C. L. (in press). Training Resource Analysis and Decision Support (TRADS) System. Proceedings of the NTSC/IST Cost Effectiveness Analysis Workshop. Held in conjunction with the 11th Interservice/Industry Training Systems Conference, Ft. Worth, TX.

Obal, J. D. (1983). Methodology for Determining Training Event Cost. Master's thesis, Monterey, CA: Naval Postgraduate School.

Orlansky, J., Knapp, M. I., & String, J. (1984). Operating Costs of Aircraft and Flight Simulators (Report No. IDA-P-1733; IDA/HQ-83-25682; SBI-AD-E500 672). Alexandria, VA: Institute for Defense Analyses.

Orlansky, J., & String, J. (1977). Cost-Effectiveness of Flight Simulators for Military Training: Volume I (Report No. P-1275-VOL-1; IDA/HQ-77-19470; AD-E500 013). Arlington, VA: Institute for Defense Analyses Science and Technology Division.

Orlansky, J., & String, J. (1977). Cost-Effectiveness of Flight Simulators for Military Training: Volume II (Report No. P-1275-VOL-2; AD-E500 014-VOL-2). Arlington, VA: Institute for Defense Analyses Science and Technology Division.

Patterson, J. F., & Adelman, L. (1981). Training Requirements and Cost Evaluation System (TRACES) for Marine Air Command and Control: Phase I [Interim report] (Report No. PR-81-26-329). McLean, VA: Decisions and Designs, Inc.

Patterson, J. F., & Adelman, L. (1982). Training Requirements and Cost Evaluation System (TRACES) for Marine Air Command and Control: Phase II [Final report] (Report No. PR-82-3-329). McLean, VA: Decisions and Designs, Inc.

Pfeiffer, M., & Gynn, S. (1986). Training Effectiveness Evaluation of the Close-In Weapon System Maintenance Trainer (Device 11G2). (Technical Report No. 86-018). Orlando FL: Naval Training System Center.

Pfeiffer, M., & Rankin, W. (1986). Training Effectiveness Evaluation of Passive Acoustic Analysis Trainer (Device 21H14). (Technical Report No. 86-019). Orlando FL: Naval Training System Center.

- Pfeiffer, M., & Scott, P. (1985). Experimental and Analytic Evaluation of the Effects of Visual and Motion Simulation in SH-3 Helicopter Training. (Technical Report No. 85-002). Orlando FL: Naval Training System Center.
- Ragusa, J. M., Barron, A. E., & Gibbons, S. (1989). User's Analysis and Evaluation of Two Computer-Based Training System Design Aids: OSBATS and ASTAR (IST-CR-89-1). Orlando, FL: Institute for Simulation and Training.
- Report on the Initial Analysis and Planning for a Cost-Effectiveness Model. (1969). El Segundo, CA: Institute for Educational Development.
- Rose, A. M. (1987). Implementation of ASTAR: Evaluation Plan for the Portable Aircraft Trainer (PAT). Washington, DC: American Institutes for Research.
- Rose, A. M., & Martin, M. F. (1988). Implementation of ASTAR: Evaluation of the Combat Talon II Maintenance Trainer. Washington, DC: American Institute for Research.
- Rose, A. M., & Martin, M. F. (1988). Forecasting Training Syatem Effectiveness: DEFT Final Report (Contract No. DAAL03-86-D-0001). Washington, DC: American Institutes for Research.
- Rose, A. M., Martin, M. F., & Wheaton, G. R. (1988). Implementation of ASTAR: Evaluation of teh Precision Gunnery Training System. Washington, DC: American Institutes for Research.
- Rose, A. M., Martin, M. F., & Wheaton, G. R. (1988). Forecasting Training System Effectiveness: Review and Analysis of ASTAR [Final Report] (Contract No. DAAL03-86-D-0001). Washington, DC: American Institutes for Research.
- Rose, A. M., & Wheaton, G. R., & Yates, L. G. (1985). Forecasting Device Effectiveness: Volume I, Issues. Washington, D.C.: American Institutes for Research.
- Rose, A. M., Martin, A. W., & Yates, L. G. (1985) Forecasting Device Effectiveness: Vol. III. Assessment of Device Effectiveness Forecasting Technique. Washington DC: American Institutes for Research.
- Rose, A. M., & Wheaton, G. R. (1984). Forecasting Device Effectiveness: Volume II, Procedures. Washington, DC: American Institutes for Research.
- Rosen, M., Berger, D. C., & Matlick, R. K. (1985). Review of Models for Cost and Training Effectiveness Analysis (CTEA) (Report No. ARI-RN-85-34). Alexandria, VA: Army Research Institue for the Behavioral and Social Sciences.

- Ross, M. C., & Yarger, G. L. (1976). A Parametric Costing Model for Flight Simulator Acquisition. Master's thesis, Air Force Institute of Technology, Wright-Patterson AFB Ohio School of Systems and Logistics.
- Rouse, W. B., Kisner, R. A., Frey, P. R., & Rouse, S. H. (1984). Method for Analytical Evaluation of Computer-Based Decision Aids [Final report] (Report No. ORNL/TM-9068). TN: Oak Ridge National Laboratory.
- Semple, C. A. (1981). Simulator Training Requirements and Effectiveness Study (STRES): Executive Summary [Final report] (Report No. AFHRL-TR-80-63). Brooks AFB, TX: Air Force Human Resources Laboratory.
- Simpson, A., & Kribs, D. (1980). Automatic Aids to Instructional Systems Development (ISD): Media Selection (Technical Rep. NAVTRAEQUIPCEN 79-C-014-1). Orlando, FL: Naval Training Systems Center.
- Skeen, J. R., & Jackson, A. E. (1981). Life Cycle Training Costs: A Literature Review [Technical note]. San Diego, CA: Navy Personnel Research and Development Center.
- Skeen, J. R., & Jackson, A. E. (1981). Life Cycle Training Cost Model [Technical note] (Report No. NPRDC-TN-81-23). San Diego, CA: Navy Personnel Research and Development Center.
- Stewart, W. B., & Kelly, D. M. (1977). Cost and Training Effectiveness Analysis for Stand-Off Target Acquisition Surveillance System (Report No.: 1979-011-1650). Annapolis, Md: ARINC Research Corp.
- Thode, D. W. F., & Walker, R. A. (1982). Establishing Relative Costs in a Complex Training System: A Cost Assessment Model and Its Application in the P-3 Fleet Readiness Squadron Aircrew Training System (Report No. NPRDC-SR-82-32). San Diego, CA: Navy Personnel Research and Development Center.
- Vaughan, D. S., Mitchell, J. L., Yadrick, R. M., Perrin, B. M., & Knight, J. R. (1989). Research and Development of the Training Decisions System (Report No. AFHRL-TR-88-50). Brooks AFB, TX: Air Force Human Resources Laboratory.
- Vecchiotti, Robert A. (1977). A Survey and Analysis of Military Computer-Based Training Systems: (A Two Part Study), Volume II, A Descriptive and Predictive Model for Evaluating Instructional Systems (Report No. MDC-E1570-VOL-2). St Louis, MO: McDonnell Douglas Astronautics Co-East.
- Weingarten, K., Hungerland, J., Brennan, M., & Allred, B. (1971). The APSTRAT Instructional Model (Report No. HUMRRO PROFESSIONAL PAPER-6-71). Alexandria, VA: Human Resources Research Organization.

Wienclaw, R. A., & Hines, F. E. (1982). A Model for Determining Cost and Training Effectiveness Tradeoffs for Training Equipment [Summary]. Proceedings of the 4th Interservice/Industry Training Equipment Conference, 1, 155-167.

Zachary, W. W. (1981). Cost Benefit Assessment of Candidate Decision Aids for Naval Air ASW (Report No. ONR-1366C). Arlington, VA: Office of Naval Research.

THIS PAGE INTENTIONALLY LEFT BLANK

