ARI Research Note 89-27

Training and Human Factors Research in Military Systems: A Final Report

Bettina A. Babbitt, Frederick A. Muckler
and Sally A. Seven
Essex Corporation

for

Contracting Officer's Representative
Charles O. Nystrom

Fort Hood Field Unit
George M. Gividen, Chief

Systems Research Laboratory
Robin L. Keesee, Director

May 1989

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

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

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

Essex Corporation

Technical review by

James H. Banks

Accession For	
NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

NOTICES

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY -- 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) --	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 89-27
6a. NAME OF PERFORMING ORGANIZATION Essex Corporation 6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute Fort Hood Field Unit
6c. ADDRESS (City, State, and ZIP Code) 741 Lakefield Road, Suite B Westlake Village, CA 91361	7b. ADDRESS (City, State, and ZIP Code) HQ TEXCOM Fort Hood, TX 76544-5065
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) PERI-S	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-83-C-0033
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 63007A; PROJECT NO. 793; TASK NO. 151; WORK UNIT ACCESSION NO. C1

11. TITLE (Include Security Classification)

Training and Human Factors Research in Military Systems: A Final Report

12. PERSONAL AUTHOR(S)

Babbitt, Bettina A.; Muckler, Frederick A.; Seven, Sally A. (Essex Corporation)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 82/01 TO 87/11	1989, May	40

16. SUPPLEMENTARY NOTATION

This research was technically monitored by Charles O. Nystrom, Norman D. Smith, and George M. Gividen, Jr., all of the ARI-Fort Hood Field Unit.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Training, Human factors, Operational test and evaluation, Combat vehicle identification (CVI), (Continued)

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

This document summarizes the technical activities conducted by personnel of the Essex Corporation in support of the work performed by the U.S. Army Research Institute, Fort Hood Field Unit, from 1 December 1982 to 30 November 1987. Three major tasks were involved: (1) Human factors technical support was provided for 26 operational tests and evaluations, including tests of vehicles; weapons systems; communications systems; and command, control, and intelligence systems. (2) Training research was conducted on combat vehicle identification (CVI) training under the Target Acquisition and Analysis Training Systems (TAATS) project, including work on advanced thermal sight training and comparative media experimental studies for CVI materials. (3) Methodologies to improve human factors operational tests and evaluations were developed.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Charles O. Nystrom	22b. TELEPHONE (Include Area Code) (817) 288-9118 22c. OFFICE SYMBOL PERI-SH

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

ARI Research Note 89-27

18. SUBJECT TERMS (Continued)

MANPRINT,
Questionnaire development,
Operator/maintainer workload assessment.

TRAINING AND HUMAN FACTORS RESEARCH IN MILITARY SYSTEMS: A FINAL REPORT

CONTENTS

LIST OF TABLES

TRAINING AND HUMAN FACTORS RESEARCH IN MILITARY SYSTEMS: A FINAL REPORT

Introduction

Support to ARI Fort Hood

On 1 December 1982, Contract No. MDA903-83-C-0033 was initiated by the Essex Corporation to provide support to the Fort Hood Field Unit of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). For the 5-year span of the contract Essex Corporation representatives provided a variety of services and products to and for ARI-Fort Hood.

According to the Statement of Work (SOW), the contractor was to "furnish the necessary personnel, materials, facilities, and other services as may be required to provide for the conduct of behavioral and social science research projects/tasks relating to training and human factors issues and problems observed or potential in military systems and organizations."

At the time that the contract began, the contractor was to aid ARI in providing support to the TRADOC (Training and Doctrine Command) Combined Arms Test Activity (TCATA). As the SOW noted, TCATA was conducting "a wide variety of operational tests and evaluations of Army materiel and tactical, training and operational concepts. Human factors evaluations are an important and integral part of each of these tests." ARI was to provide support to TCATA with respect to the human factors aspects of the tests, and Essex Corporation was contracted to help ARI-Fort Hood in providing that support.

Program Task Structure

Four specific tasks were called out in the original statement of work:

- Research on Human Factors and Training Issues in TCATA Operational Tests and Evaluations
- Research on Target Acquisition and Analysis Training Systems (TAATS)
- Human Resources Test and Evaluation System (HRTES)
- Questionnaire Development Research

Each of these four tasks was developed in considerable detail in the SOW. In detailing the first task listed above, the SOW named three specific tests that were to be covered: Joint Tactical Fusion Test Bed, M1E1 Tank, and Remotely Piloted Vehicle (RPV). Although only these three tests were called out in the SOW, it was anticipated that a number of other tests would be covered under this contract during the five years that it was to be in force. As the SOW noted, additional tests for the earlier years of the contract, as well as tests for later contract years, could not be specified at the time of contract award because the specific concepts and equipment to be tested were not yet firm.

In June 1983, ARI and USAOTEA (United States Army Operational Test and Evaluation Agency) signed a Letter of Agreement which later enlarged the scope of the tasks Essex would perform under its contract with ARI. OTEA plays the major role in the operational test and evaluation of Army systems and equipment as they are being procured and developed. As with TCATA, ARI help was to be provided to OTEA on human factors aspects of test performance. According to the agreement reached with OTEA, ARI and its contractor were to provide human factors, safety, and training evaluations of systems being acquired by the Army (AMC, typically) at the time OTEA conducted its operational test and evaluation of the systems. It was agreed that Contract No. MDA903-83-C-0033 would be modified to support and implement the ARI-OTEA agreement.

In addition to the expansion of the first task, it was recognized that the third and fourth tasks listed in the original SOW were related, in that they were instances of methodological concerns. To clarify the focus of the contract and to avoid a proliferation of separate tasks, the SOW was revised in November 1983 and the revision was incorporated as a contract modification (P00007) in April 1984.

The revision of the SOW accomplished two things. First, it recognized the expanded scope of individual test coverage as one focus of the contract. Second, it clarified the fact that an underlying and essential part of the contract was a concern for improving the methodology available to and applied in the operational test and evaluation (OT&E) environment. The third and fourth tasks listed in the original SOW (HRTES and Questionnaire Research) had certainly dealt with methodological issues, but additional methodological questions relevant to OT&E could be expected to arise (and did) during the five years the contract was to run.

The revised work statement specified three tasks, expanding human factors and training in OT&E, leaving the second task (TAATS) intact, and creating a generic methodology task that would encompass a number of specific projects. It was hoped that such projects would lead to improvements in system assessment and ultimately in the quality of the systems fielded at the end of the procurement process. The November 1983 SOW listed these three tasks:

- Research on Human Factors and Training Issues in Operational Tests and Evaluations

- Human Factors Research on Target Acquisition and Analysis Training Systems (TAATS)

- Improved Methodologies of Assessment of New Systems in an Operational Test Environment

The new three-task breakdown provided the framework under which Essex Corporation conducted a variety of endeavors in support of the mission of the Fort Hood Field Unit of ARI. Throughout the life of the contract, contractor personnel were on-site at Fort Hood, TX to facilitate performance of contract activities. Essex representatives were housed in Government-provided workspace. In addition to the work conducted at Fort Hood, activities in support of this contract were carried out at a wide variety of sites. Task 1 locations were particularly varied since they were dependent on the test sites chosen for the specific equipment being tested.

Essex personnel at the Fort Hood facility were not the only ones involved in this ARI contract. Including those at Fort Hood, a total of 33 technical personnel participated in one or more of the three tasks subsumed under this contract. Participants are identified in the discussion of each specific task.

This report is the final summary of activities conducted during the period from 1 December 1982 to 30 November 1987 under Contract No. MDA903-83-C-0033. This summary is organized in keeping with the three-task structure of the November 1983 SOW. This report is the final deliverable due under the contract and it references other deliverables produced during the 5-year span of the contract.

Task 1: Human Factors Support to Operational Test and Evaluation

Objective of Task 1

As stated in the Introduction, the first of the three basic contract tasks was research on human factors and training issues in operational tests and evaluations. Since 1971, the Army Research Institute (ARI) has provided a Fort Hood Field Unit principally to support operational tests and evaluations (OTE) performed at the TRADOC (Training and Doctrine Command) Combined Arms Test Activity (TCATA) at Fort Hood. The evaluation plan for virtually every OTE has contained a Human Factors issue, less frequently a training issue. These issues, have been combined with Manpower, Personnel, Safety, and Health Hazards issues to form the MANPRINT issue in very recent years. ARI-Fort Hood has provided support to TCATA tests, and the contract in turn was designed to provide support to ARI-Fort Hood human factors experts. OTE support was the major single task performed over the period of the contract.

A History of Task 1: 1982-1987

Over the 5-year period of the contract, twenty-six (26) operational tests and evaluation were supported through ARI-Fort Hood. These tests are listed in Table 1. As may be seen there, they have been divided into the four categories of: vehicles; weapons and weapons subsystems; communications systems; and command, control, and intelligence.

A number of changes occurred over the period of the contract which had significant impacts on the execution of the tests:

Extension beyond TCATA

It was originally expected that contractor support would be provided just to TCATA, but subsequent events expanded support to include testing done by the U.S. Army Operational Test and Evaluation Agency (OTEA). In the case of the M109 HELP system, human factors support was provided to the U.S. Army Field Artillery Board. Support of OTEA tests was provided by ARI in-house personnel for over two years prior to the use of contractor support. The reduction of in-house staff available for OTEA test support, combined with OTEA's request for increased HF and safety evalation, led to the arrangement whereby ARI procured and monitored contractor efforts, while OTEA provided the required funding.

Extension beyond HF and Safety

Originally, the technical focus of each test was on human factors engineering of the system, coupled with concern for safety-related items; in some cases, training implications were also considered. With the advent of the Army-wide MANPRINT Initiative (Army Regulation 602-2, 18 May 1987), the technical emphases of the tests were expanded to cover all of the six MANPRINT categories: human factors, manpower, personnel, training, safety, and health hazards. This expansion imposed a significantly greater burden on the test teams in terms of resource requirements and test conditions for deriving desired MANPRINT data and information.

Table 1

Human Factors (MANPRINT) Tests and Evaluations Conducted for the Army Research Institute, 1983-1987, Contract No. MDA903-83-C-0033

System Type and Acronym	System Name
Vehicles	
AMV	Armored Maintenance Vehicle
M9 ACE	Armored Combat Earthmover
RPV	Remotely Piloted Vehicle (Aquila)
HMMWV	High Mobility Multipurpose Wheeled Vehicle
LAVM/RV	Light Armored Vehicle, Maintenance Recovery Vehicle
FIST-V	Fire Support Team Vehicle
Weapons and Subsystems	
M1, M1E1, M1A1	Abrams Main Battle Tank
M2FV, M3FV	Bradley Fighting Vehicle
M109 HELP	Howitzer Extended Life Program
STE-M1	Abrams Test and Bite Maintenance
SGT YORK	Division Air Defense Gun
PATRIOT	Air Defense Missile System
Communications Systems	
SCOTT	Single Channel Tactical Terminal
SST	Single Subscriber Terminal
MSE	Mobile Subscriber Communications
TC3S	Tactical Command, Control, Communication System
RECS	Rear Echelon Communication System
TACJAM	AN-MLQ34 Electronic Countermeasures Set
SINCGARS	Single Channel Ground Airborne Radio System
AN/TRC 170(V)	Digital Atmospheric Troposcatter Relay
Command, Control, and Intelligence	
NAVSTAR GPS	Navigation Global Positioning System
JTIDS	Joint Tactical Information Distribution System
JINTACCS	Joint Operability Tactical C&C System
JTFTB	Joint Tactical Fusion Test Bed
ASAS	All-Source Analysis System
ARTBASS	Army Training Battle Simulation System

Performance Data

It was the understandable desire of ARI to obtain more hard performance data from the tests relative to the amount of subjective, participant or expert observer data collected via questionnaires. That is, ARI required that more objective data on the soldier-machine interface be collected to identify and substantiate soldier performance and the impact of soldier performance on system performance. While this goal was not new to human factors experts in operational test and evaluation, it placed a number of demands on measurement resources which were very difficult to meet. Be that as it may, eventually a great deal of objective performance data was obtained in, for example, the Sgt York and RPV tests.

Some of these tests were significant technical milestones for the contract team including:

NAVSTAR Global Positioning Satellite: Human factors scoring conference
Light Armored Vehicle Maintenance/Repair Vehicle: Extensive photographic documentation
M1E1 Abrams Tank: Soldier performance side test
M9 Armored Combat Earthmover: Comprehensive data collection plan and technical video report
SGT YORK and Remotely Piloted Vehicle: Extensive individual and crew performance measures

In short, for these tests, additional technical approaches were implemented beyond the previously customary questionnaire and interview techniques.

A fundamental reality of all operational tests is insufficient time and resources to collect all the data that each test specialty could put to good use. Apparently this will always be the case; it is imperative that test specialists be trained to understand and cope with that fact; and it is a pleasant surprise when more extensive technical work can be performed. When such an expansion is possible, much better data can be collected.

It is difficult to estimate the degree of "success" that was achieved on the 26 tests listed in Table 1. In every case, there was identification of human factors/MANPRINT deficiencies in the specific system. It was rewarding to know that subsequent action was taken to correct some of the deficiencies. This was the case, for example, with the M109 HELP project where the system designer was aware of the system deficiencies and took design steps to correct them. Briefly, a serious technical problem was identified in the TACJAM and was fixed by the designer. Past and present tests of the Abrams Main Battle Tank showed that many serious deficiencies from the soldier-system standpoint were identified and had been corrected in new versions of the tank.

But, unfortunately, other problems surfaced again and again. For tracked vehicles such as the Bradley M1 and M2, and the Sgt York, the driver's station is too small. The use of MOPP face masks is simply not compatible with the use of optical devices. Storage space is consistently inadequate for manuals, MOPP gear, and other personnel equipments. It is suggested that there be periodic conferences of human factors/MANPRINT operational test specialists to identify and record those technical problems that exist across systems and persist through systems.

Principal Products

In participating in any operational test and evaluation, the human factors (MANPRINT) practitioneer has four general responsibilities:

- Generation of the human factors/MANPRINT test plan that will provide specific human factors/MANPRINT data as well as be consistent with the overall test plan;
- Participation in the execution of the test itself;
- Generation of appropriate results for the test report to be submitted by the testing agency (often within 30 days after the test); and
- Where appropriate, generation of a separate technical report on the test.

Essex test personnel completed all of the first three steps for all of the tests shown in Table 1. The following are test plans and technical reports prepared and submitted for various tests:

Bowser, S. E., Lyons, L. E., and Heuckeroth, O. (1987, April). MANPRINT test report of the follow-on operational test and evaluation of the AN/TRC -170(V).

Cotton, J. C. (1985, February). RPV: Human factors evaluation plan.

Cotton, J. C. (1985, April). Mobile subscriber equipment (MSE): Summary report on the system operating suitability and human factors assessment of the GTE/RITA and Rockwell alternate MSE systems.

Cotton, J. C., & Bowser, S. E. (1984, August). NAVSTAR-GPS: Human factors evaluation plan.

Kubala, A. L. (1987, August). A preliminary MANPRINT evaluation of the All Source Analysis System (ASAS).

Krohn, D. A. (1987, June). High Mobility Multipurpose Wheeled Vehicle - Heavy Variant (HMMWV-HV): Assessment of MANPRINT concerns for predecessor vehicles prior to HMMWV-HV operational assessment.

Krohn, D. A., & Spiegel, D. K. (1987, October). HMMWV OTEA Test Report: Appendix B. MANPRINT Assessment.

Krohn, G. S. (1985, February). Human factors assessment: M1 combat tank simplified test equipment (STE-M1) and built-in test equipment (BITE).

Krohn, G. S. (1986, February). Human factors assessment: M9 Armored Combat Earthmover (ACE).

Krohn, G. S. (1986, December). Human factors assessment: Light Armored Vehicle, Maintenance Recovery Vehicle (LAVM/RV) (ARI Research Report 1434).

Krohn, G. S., Kubala, A. L., & Earl, W. K. (1985, February). Human factors assessment: M1E1 tank.

Krohn, G. S., & Lyons, L. E. (1987, January). MANPRINT assessment plan and data collection materials for the M1A1 Abrams tank follow-on evaluation.

Krohn, G. S., McFarling, L. H., & Bowser, S. E. (1986, May). U.S. Army Remotely Piloted Vehicle (RPV): Human factors engineering analysis plan and data collection materials.

Krohn, G. S., Spiegel, D. K., & Kelley, G. (1987, July). RPV OTEA test report: Appendix B. MANPRINT assessment.

Lyons, L. E. (1983, August). Human factors evaluation: AN-MLQ 34 electronic countermeasures set (TACJAM).

Lyons, L. E. (1986, October). Human factors support to the M109E4 (HELP) howitzer OT II. Fort Sill, OK: U.S.A. Field Artillery Board.

Lyons, L. E. (1986, November). Human factors support to the Armored Maintenance Vehicle - powerpack transport trailer concept evaluation program (AMV/PTT CEP).

Lyons, L. E., & Nystrom, C. O. (1984, March). Human factors evaluation: M2/M3 "Bradley" Fighting Vehicle.

Pieper, W. J. (1986, September). High Mobility Multipurpose Wheeled Vehicle - Heavy Variant (HMMWV-HV): Human factors engineering analysis plan.

Pieper, W. J., & Avery, L. W. (1987, September). JTIDS investigative operational assessment MANPRINT.

Shaw, B. E. (1985, February). Single-Subscriber Terminal (SST): Human factors evaluation plan.

Shaw, B. E., & Pieper, W. J. (1985, October). Human factors evaluation findings: AN/UGC 137 A(V)2 single subscriber terminal.

Videotape Summaries

In some of the tests, videotape recording was used to photograph certain operator/maintainer task performance. The videotape raw data were used to prepare summaries of the tests in two cases, the M9 ACE and the RPV.

Lessons Learned

Over the period of contract performance, approximately two-thirds of the contract effort was given to operational test and evaluation. At any given time, technical personnel were participating in from four to eight tests at various stages in the test execution cycle, ranging from initial test planning to final test report writing. From participation in the 26 tests, a number of lessons were learned or, perhaps better, relearned since they will not be new to test and evaluation specialists.

Participation

The experience of participating in operational tests and evaluations is an opportunity for involvement in new Army systems and advanced Army soldier-system technology. For the first time, new technology is being exercised in an approximation of the operational environment, and the "good" and the "bad" of the new technology become apparent. Unsolved problems reappear. Promising new solutions emerge. It is a chance to glimpse the Army of the future; it is an exciting experience for most human factors specialists.

Soldier-System Interface

In every system, the soldier-system interface is essential. There is no such thing as a completely "automated" system; all current and projected Army systems will continue to require critical and good performance from the soldier. Therefore, it is imperative that the soldier-machine interface be carefully checked in the operational test. There must be human factors/MANPRINT test specialists to identify human factors/MANPRINT deficiencies, establish their importance for system performance, and assist in correcting them.

Limitations

As noted above, for every operational test, time is short and resources limited. And the behavior to be tested and measured is very complex. This means that very careful planning must be done to collect those data that are essential and to minimize the collection of less useful data. This requires a great deal of skill by the test participants. Often, resources are not available to collect exactly the data that are needed and substitute measures must be found. Therein lie many technical challenges.

Covering All MANPRINT Domains

One such challenge is testing adequately for all categories of the MANPRINT Initiative: human factors, manpower, personnel, training, safety, and health hazards. The technology for testing human factors, safety, and health hazards seems to be fairly well advanced and usable. But there are many questions as to appropriate testing for manpower, personnel, and training variables.

In the first place, what are the appropriate questions for manpower, personnel, and training? For example, estimates will have been made for the manpower requirements of the system under test for the entire life cycle of the system. What data, if any, collected during test can verify or deny the validity of those estimates? Or, with respect to personnel dimensions, what test data can be collected to demonstrate whether or not the system can be operated by Category III or Category IV Army personnel?

One problem here is that test subject sample sizes are usually very small; often as few as 3 to 9 soldiers may be used in the test. While a great deal will be known about those particular soldiers, it is not easy to generalize to the overall Army population of potential system users.

Training presents two kinds of problems in operational testing. First is the necessity to train the test sample of soldiers to use the system sufficiently so that the system can be exercised and evaluated. In most of the tests in which we have participated, the test soldier-users were not adequately trained for the system test, and the results were confused by the question of whether poor performance was due to inadequate training or to a poor soldier-machine interface. Thus, performance may have been poor simply because the operators, during testing, were still learning how to perform their tasks. The second problem is using the operational test and evaluation data to make estimates about future system training requirements or the adequacy of the training designed for the system. Some experts feel that there is insufficient data to do either.

Finally, even though the Army wants MANPRINT data and information, specific constraints on individual projects may preclude the measurement of one or more of the MANPRINT dimensions.

Chronic problems

In all operational tests, there are a set of chronic problems that recur:

- Budgets are inadequate.
- Schedules are too short.
- Planning is made even more difficult by constant changes.
- Field conditions are rigorous, demanding, and sometimes hazardous.
- Deficiencies, once found, are not easy to get corrected.

From one point of view, the last issue is not relevant to operational test and is, in fact, a "problem" to be faced by some other part of the Army system. The operational test identifies deficiencies and evaluates their potential consequences — and that is all. Yet it is not clear who in the system is responsible for fixing human factors/MANPRINT operational system problems or who has the authority to see that they are fixed. Surely the purpose of OT&E is to improve systems, not just to know what is wrong with them.

Improving test methodology

A particular operational test is not the place to invent better test methodology. There is neither time nor resources. Sometimes advances are made, but the operational test specialists from human factors or MANPRINT must go with the technology that is available. There is a need, therefore, for improvement in MANPRINT methods that can be applied in operational test and evaluation. Some of the resources of the current contract were assigned to that goal, and the results are discussed in the section on Task 3.

Program Personnel

The following individuals participated in and contributed to the tests conducted under Task 1:

Mr. Larry W. Avery
Dr. Bettina A. Babbitt
Dr. Samuel E. Bowser
Mr. John C. Cotton
Mr. Charles H. DeBow
Mr. Tracy Edaburn
Mr. Richard Hiss
Mr. Gene Kelley
Mr. Gregory S. Krohn
Mr. Daniel A. Krohn
Dr. Albert L. Kubala
Mr. Lawrence E. Lyons

Dr. Leslie McFarling
Mr. Carl Mortenson
Dr. Frederick A. Muckler
Mr. Douglass R. Nicklas
Dr. Richard F. Pain
Mr. William Pieper
Dr. Charles R. Sawyer
Mr. Brian E. Shaw
Dr. Douglas K. Spiegel
Mr. Mannon Thomason
Mr. William L. Warnick
Mr. Charles L. Wright

When assigned to specific operational tests, these individuals were promised three things: a tremendous technical challenge, technical opportunity, and demanding environments. For the most part, they got all three.

Task 2: Target Acquisition and Analysis Training Systems (TAATS)

Objective of Task 2

The second major task of the contract was research on target acquisition and analysis training systems (TAATS). Since the beginning of warfare, the identification of friend or foe has been a consistent battlefield problem. Under stress, with incomplete information, poor visibility, and a host of other concerns, friends are all too likely to be fired upon. There were recent examples of such mistakes during battles in the Faukland Islands and Lebanon.

To reduce this problem, for years there have been training programs for combat vehicle identification (CVI) so that both friends and foes can be detected and recognized. Both the Systems and the Training Research Laboratories of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) realized over a decade ago that research in the area of CVI could produce major operational training benefits. While there have been many efforts to produce automated identification for friend and foe (IFF), the fact still remains that on the battlefield, the soldier must do unassisted CVI. If he is to do it well, he must be trained for the task.

Early experiments by ARI personnel focused on the need for CVI at distances for long-range engagement. They were concerned with optical devices such as binoculars and gunsights, none of which had significantly positive effects on CVI. Indeed, CVI performance was found to be poor. As a result, several CVI training packages were developed to meet an immediate need in the recognition and identification (R&I) area.

The first prototype training program was the Basic Combat Vehicle Identification (CVI) Training Program, developed and offered to 22 operational units. The prototype program achieved these designed objectives:

- Train soldiers to recognize primarily those cues important for vehicle identification at realistic combat ranges
- Keep training simple with a minimum of support materials
- Be modular in design and usable in short training periods
- Be adaptable for use in simulation of any potential distance requirement
- Produce high levels of motivation and learning in a short training period
- Standardize training for recognition and identification in the Army
- Provide an ongoing measure for evaluating R&I training skills

These objectives were kept for the entire program, from 1979 through 1987.

Following the development and test of the Basic CVI program, a number of further developments were made:

- An Advanced CVI training program was developed that included target masking.

- In 1981, a set of flashcards was designed to provide the soldier with a method for reviewing the R&I training he had received in the Basic CVI program. There were two views of each of 30 vehicles on 60 cards with identification cues on the back, plus an overall instruction card. Upon empirical evaluation, the flashcards were shown to be a useful adjunct to training.

- A training program for operators in the Operational Test II of the Remotely Piloted Vehicle (RPV) was completed in 1982.

- A Basic Thermal Combat Vehicle Identification Training Program (TCVI) was developed and tested in 1982. This application was to enhance the effectiveness of the use of the many thermal sights operational in the Army, particularly in self-propelled vehicles.

In each case, the developed program was subjected to empirical field test to evaluate training effectiveness. Significant improvements in R&I performance were obtained in each case.

A History of Task 2: 1982-1987

Many tasks were performed on the TAATS portion of the contract from 1 December 1982 through February 1987. They can be divided into three sets: (1) continued enhancement of the basic stimulus library for CVI training materials, (2) new media developments, and (3) experiments on the effectiveness of CVI training.

CVI Stimulus Library

During the development of the CVI training programs, there was a constant requirement for appropriate visual stimuli of friendly and hostile vehicles. For training purposes, views of the vehicles at varying distances and varying aspect angles were desired. It was exceedingly difficult, often not possible, to get such imagery for vehicles of other nations, not just Warsaw Pact nations.

In the absence of actual views, a variety of techniques were used to develop simulated views by either physical simulation and/or graphic simulation. Both methods appeared to be effective. In the former, small scale models were obtained or made and then photographed appropriately. A very high level of realism could be obtained. Further, the vehicles could be imbedded in a variety of operational backgrounds using varying degrees of masking. For future CVI training, this technique was found to be very realistic and cost effective.

New Media Developments

In order to keep production costs low and to make use of audiovisual devices that were already in the Army training device inventory, the Basic CVI was designed around the use of 35-mm slides and the 35-mm carousel slide projector. Once the training program was distributed to military units, many units reported that they "had problems with the slides." They were easily lost or damaged. Even though each slide was coded, they got mixed up in the slide trays.

It appeared that the use of 35-mm filmstrips would overcome many of the problems associated with slides. The entire Basic CVI program was converted to a filmstrip format. All the necessary equipment to teach the program was packaged in a container no bigger than an attache case. Two prototypes were delivered to ARI.

Experiments on CVI Effectiveness

Under the direction of Dr. Norman D. Smith of ARI, who functioned as Assistant Contracting Officer's Technical Representative for Task 2, contractor personnel participated in four major experimental studies on CVI effectiveness.

Comparison of image quality of three presentation media. This study was an evaluation of the effects of quality of visual imagery as presented by the various media (35-mm slides, 8-mm Bessler, and videotape) on soldier's recognition and identification performance. Three training modules extracted from the Basic CVI training program were used for this evaluation.

There were no significant differences in cognitive and performance measures among any of the three media systems. The six military/civilian trainees who participated in the research were asked to rate the following: quality of imagery, training effectiveness, ease of use by the trainees, suitability for soldiers of differing abilities, and level of soldier interest. The videotape and 35-mm slide systems received essentially the same high positive ratings, while the Bessler 8-mm system was rated somewhat less positively.

Comparisons of performance before and after training clearly indicated that the training concepts embodied in the Basic CVI may be taught on all three of these systems.

Vehicle identification performance using thermal images. Following completion of the Basic CVI program for photopic imagery, high priority was given to the development of a thermal CVI program (now designated by the Army as GTA 17-2-10). When IR sensing systems were first added to such weapons systems as the TOW in the 1970s, user comments led to a decision to include a polarity switch that would permit displaying both black and white hot images. In our development efforts, black hot images were selected since white hot images seemed to bloom or blossom, especially at extended ranges, and therefore appeared to make distinguishing among vehicles more difficult. This research was designed to provide specific information about the role of polarity setting during target acquisition and identification.

The major objectives of the research were to determine (1) overall vehicle identification performance differences between black and white thermal image polarity settings; (2) how vehicle identification performance is affected by black and white thermal image settings at different ranges; (3) the general pattern of identification to thermal images (black and white combined); and (4) how vehicle identification performance for each vehicle varies depending on the image polarity setting used.

Major findings of this research indicate that vehicle identification is generally better to a black hot image, but several vehicles are apparently easier to identify from a white hot image. Whereas earlier research has shown identification performance to photopic images is generally comparable over a 1,000 to 3,000 meter range, identification of thermal images for many vehicles degrades rapidly beyond 1,200 meters when either white or black hot imagery is used.

Conclusions:

- Identification performance is generally better using black hot thermal imagery, which is the form of thermal imagery preferred by participating soldiers.

- The fact that identification performance for several vehicles is superior with a white hot setting suggests that further research would be instructive on which vehicles have characteristics that are revealed best with high contrast black hot setting or a high intensity white. Operationally, should CVI not be accomplished in a timely fashion with the black hot polarity setting, attempts should be made to adjust contrast and brightness settings to highlight specific differentiating features. If still unsuccessful, a switch to white hot polarity should be made and followed with the adjustment procedures used with black hot.

- Compared with earlier research using photopic imagery, identification performance to thermal images of many vehicles degrades sharply beyond 1,200 meters.

- Compared to range and vehicle type, individually and in combination, image polarity appears relatively less important to vehicle identification of thermal images.

Comparison of two training media using the Basic CVI Training Program. A 35-mm Kodak slide projector system was initially used as the medium for the CVI training program because it was available throughout the Army supply system. Several problems, however, were reported with the use and maintenance of the CVI slide system. The ARI-Fort Hood Field Unit was, as a consequence, tasked to develop and evaluate an alternative CVI filmstrip system (described above).

The evaluation found that the audio filmstrip system was a better all-around training tool than was the slide system. The soldiers who were trained with the audio filmstrip system first showed significantly higher identification performance than did the soldiers who were trained with the slide system first. Furthermore, on the second day of the training period, the soldiers who then received audio filmstrip system training produced significantly higher identification performance than they had on the slide system. On the other hand, the soldiers who received slide system training on the second day showed no significant increase in identification performance over their earlier performance on the audio filmstrip system.

Examination of the questionnaire responses found that soldiers preferred the audio filmstrip system over the slide system both for its ease of use and maintenance, and as a system on which they would prefer to receive other training.

Conclusions:

- The audio filmstrip system trained combat arms soldiers to significantly higher combat vehicle identification performance than did the slide system.

- The soldiers who participated in the research rated the audio filmstrip system significantly easier to use and maintain than the slide system.

- The junior enlisted soldiers reported that they would prefer to receive future training with the audio filmstrip system. Furthermore, the NCOs reported that they would prefer to give future training with the audio filmstrip system.

Effects of motion on performance in the CVI Training Program. In training, motion is often thought to be an important ingredient. Many feel it adds realism and increases soldier motivation to learn as well as providing additional information about the vehicles in the form of cues which facilitate learning. The purpose of the fourth experimental study was to test the validity of the belief that adding motion to vehicles in the Basic CVI training program improved performance.

Data from the 85th Army Reserve Division (Training), Arlington Heights, IL, were used to evaluate the effects of motion during training and for overnight retention. Soldiers (N = 120) were assigned to one of four conditions: circular motion, rotational motion, straight-line motion, or static (no motion). The training medium was videotape. Three modules comprising a total of 15 vehicles from the Basic CVI training program were used.

Motion did not contribute to improved performance in initial learning or in the overnight retention. Groups trained to recognize and identify combat vehicles in simulated motion (models moving on a table top) learned them no better than groups trained on static vehicles. With repeated training, however, especially for soldiers who experience difficulty in acquiring R&I skills, a small benefit from training involving motion appeared.

Conclusions:

- Motion (after repeated training) provides a small effect but does not appear to be an essential ingredient in training ground-to-ground vehicle R&I using the Basic CVI training program.

- Short-term R&I retention is not improved when motion is included in the training.

- Soldiers who experience difficulty in acquiring R&I skills may benefit slightly from repeated training with motion; however, cost-effectiveness considerations make this solution, at best, questionable.

- The introduction of motion to CVI training is not cost effective.

Implications of the studies

The results of these four studies have implications beyond the R&I context, and provide interesting evidence for general training system development issues. For example, one of the most important decisions in all of training system development is the choice of appropriate training media. In this task, there was a direct comparison of three media, and there was no differential training effect (35-mm slides, 8-mm Bessler, and videotape). On the other hand, a filmstrip system was found to be superior to 35-mm slides in terms of ease of operation and maintenance. These findings should be included in the general literature on media selection.

A second issue of tremendous importance in training system development is fidelity of simulation, i.e., how close to the actual stimuli must the training material be? The lack of any significant effect due to motion in the vehicles reflects a common finding that low levels of fidelity can achieve high levels of training. In this case, at least, the vehicles do not have to appear to move in order to achieve high levels of CVI performance. The cost-effectiveness implications should be obvious.

Finally, outside of training issues, the question of black hot and white hot thermal images concerns basic soldier task performance using the thermal sight. R&I is a basic task for the soldier, and the thermal sight should contribute to that task. The data found here suggest that beyond 1,200 meters, R&I performance is poor. The use of the polarity switch and the alternative images may help, but even that is not clear. What the results do suggest is taking a closer look at the performance obtained from thermal sights. Parenthetically, in preparing for these studies, it was found that soldiers could not use the thermal sights well, that they needed training on the use of sight controls, and that it later proved necessary for us to prepare a videotape training package that really did not concern R&I so much as it concerned how to use the device.

Principal Products

There were three principal classes of products from the TAATS task: (1) improved imagery for stimulus material in the training packages, (2) technical reports, and (3) videotapes on specific topics associated with the TAATS project.

Imagery

Appropriate stimulus imagery is essential for any training packages in combat vehicle identification. Over the contract period, much effort was given to updating and improving the vehicle imagery available. This included:

1. Continuing updates of the pictures of foreign and American tracked and wheeled vehicles, both friend and hostile.

2. Imagery of friendly and hostile vehicles obtained through the use of thermal sights.

3. Black hot versus white hot thermal images of friendly and hostile vehicles.

4. Computer-enhanced images.

For future work in combat vehicle identification training, the file of stimulus material at ARI-Fort Hood may be of value.

Technical Reports

In addition to supporting ARI technical reports of the experimental work in TAATS, the following documents were completed:

Kubala, A. L. (1986, February). Potential predictors of target acquisition performance by gunners: A literature review.

Kubala, A. L. (1986, March). Approaches to the development of thermal imagery for recognition and identification training programs.

Warnick, W. L. (1986, July). Target Acquisition and Analysis Training System (TAATS): A training systems approach to target acquisition and analysis.

The following reports were produced jointly with ARI personnel:

Heuckeroth, O. H., Smith, N. D., Warnick, W. L., Kubala, A. L., Lyons, L. E., & Maxey, J. M. (In press). Target acquisition and analysis training system: A preliminary study of vehicle identification performance to black hot and white hot thermal images.

Shope, G. L., Smith, N. D., Heuckeroth, O. H., Warnick, W. L., & Essig, S. S. (1984). Evaluation of an advanced combat vehicle identification (CVI) training program (masking): A new approach to target acquisition training (ARI Research Report 1368).

Smith, N. D., Heuckeroth, O. H., Shope, G. L., Warnick, W. L., & Essig, S. S. (In press). Target acquisition and analysis training system: Effects of motion on performance in the combat vehicle identification (CVI) training program.

Smith, N. D., Shope, G. L., Heuckeroth, O. H., Nystrom, C. O., Betts, S. W., Warnick, W. L., & Essig, S. S. (In review). Target acquisition and analysis training system: Comparison of image quality of three presentation media. (ARI TR 622).

Smith, N. D., Shope, G. L., Heuckeroth, O. H., Warnick, W. L., & Essig, S. S. (1983). Target acquisition and analysis training system: An evaluation of the basic thermal combat vehicle identification (TCVI) training program (ARI Research Report 1378).

Smith, N. D., & Warnick, W. L. (1983, August). TAATS: A training systems approach to target acquisition and analysis. Army-wide Vehicle Recognition/Target Acquisition and Analysis Training System Seminar, 30-31 August 1983, Fort Leavenworth, KS.

Videotapes

Over the course of the contract, six videotape presentations were prepared:

- A videotape presentation was prepared of the Basic CVI program for showing both as a demonstration and as a training device (1984).

- Thermal imagery was prepared and processed, using videotapes as the storage media (1983-1984).

- A videotape was prepared on the management of CVI for service schools (1985).

- A videotape entitled "Training for Combat" was prepared as an update for the CVI and TAATS training packages (1986).

- A videotape derived from CVI work on the M1 was prepared entitled "M1 Abrams Tank Thermal Sight Adjustment Program" which was a training presentation designed for better use of the thermal sight as well as better combat vehicle identification.

- A videotape on TAATS was prepared for a NATO conference presentation.

Video tape presentations were also used in the experiments concerned with comparison of three presentation media, and the effects of image motion on CVI training performance.

Future Developments

The ARI-Fort Hood TAATS program was terminated in 1987, and Essex work was closed out, with the exception of contributions made to the preparation of a final summary report on the entire program by Dr. Norman D. Smith of ARI, who was the TAATS Research Task Principal Investigator.

The battlefield problem of recognition and identification of friend and foe will continue, and a need for training programs for R&I and CVI will also continue. The value of such training seems particularly high. For the Airland Battle 2000, friendly force attrition is a critical matter and should not be made worse by the inadvertent destruction of friendly forces. Further, it has been shown in the history of the CVI and TAATS program that very effective training can be achieved at very low cost. Indeed, saving one Abrams Main Battle Tank from friendly destruction would pay for the entire cost of the CVI and TAATS program from inception.

It is to be hoped, then, that CVI training will continue to be sponsored and funded by one or more agencies of the Army. If so, these guidelines are offered based on the history of the program:

- Soldier R&I performance is normally very low, and continuing operational training is essential.

- Very low-fidelity, low-cost training devices serve very well to achieve high R&I performance.

- Acquisition and maintenance (to include updating) of an extensive friend-foe visual imagery file is essential and this is not easy to do. But it is a problem that can be overcome by small-scale model simulation and effective photography and videography.

- Careful design of the operational training devices for CVI is essential for usability, logistics, and reliability. Simple, reliable, and durable devices are needed, and the experience from this program demonstrates that they are effective for training.

The work on thermal sights and soldier R&I performance using the sights raised many questions about soldier visual performance and the basic effectiveness of the sights. We observed the soldiers having difficulty using the sights. This suggests that additional human factors engineering attention might be given to the whole domain of thermal sights and soldier visual performance with them.

Program Personnel

The following individuals participated in and contributed significantly to Task 2:

Mr. Randy C. Bruce	Mr. Carl Mortenson
Ms. Frances S. Essig	Dr. Frederick A. Muckler
Mr. Stephen S. Essig	Mr. David Pigeon
Ms. Teri Gray	Mr. F. L. Rozinka
Mr. John R. Haefner	Ms. Deborah Sandler
Mr. Richard Hatch	Mr. David Sandlin
Ms. Karen Hughes	Ms. Marcia Shanahan
Dr. Albert L. Kubala	Ms. Sally Spitzer
Ms. Johanna Lack	Mr. Michael Sterns
Mr. Lawrence E. Lyons	Ms. Kim M. Vetter
Mr. James M. Maxey	Mr. William L. Warnick
Mr. Henry J. Miklaski	Mr. Charles L. Wright

The Task 2 Team Leader was Mr. William L. Warnick. Principal technical direction was given by Dr. Albert L. Kubala.

Task 3: Improving Human Factors (MANPRINT) Test and Evaluation Methodology

Objective of Task 3

Task 3 was concerned with a variety of methodological issues relevant to human factors evaluation in operational settings. Its objective was to improve methods of assessing the human factors aspects of new and developing systems in an operational test environment.

During the life of this contract, the major responsibility for evaluating the overall operational effectiveness of major weapon systems that were underdevelopment belonged to the United States Army Operational Test and Evaluation Agency (USAOTEA). A critical aspect of this responsibility was assuring thatthe human resources elements of the overall system were adequately recognized and accommodated as part of the evaluation process. ARI was concerned with the human factors and training aspects of system evaluation and recognized that in order to evaluate human resources aspects of a system effectively, it was important to have available information on appropriate evaluation methods, including knowing what questions to ask at a given phase in evaluation. These concerns were taken as the starting point for a series of projects that addressed methodological issues in operational test and evaluation (OT&E) and that together made up Task 3.

A History of Task 3: 1982-1987

The first project initiated under Task 3 was an evaluation of the Human Resources Test and Evaluation System (HRTES) Handbook. Subsequently, and continuing through 1987, several other research projects were conducted on topics dealing with techniques such as questionnaire construction, interviewing, and workload assessment. Products were developed to improve future test planning and to answer methodological questions concerned with enhancing the state of the art in MANPRINT assessment in the context of operational test and evaluation. A description of each of the Task 3 projects and the products that resulted from them follows.

Human Resources Test and Evaluation System (HRTES)

To assist those who conduct human factors evaluation of military equipment, ARI had contracted with Perceptronics for the preparation of a handbook to provide guidance on human factors assessment in OT&E. Perceptronics produced a two-volume handbook called Human Resources Test and Evaluation System. Because Volumes I and II were extremely detailed and lengthy, ARI and OTEA agreed upon (and jointly funded) a contract modification in which Perceptronics produced a shorter version of the handbook. The shorter version was known as "Quick HRTES." It, too, was in two volumes. Volume I, "Test Procedures," described the procedures that should be followed in conducting an evaluation. Volume II, "Supplement," contained detailed descriptions of test procedures and methods as well as forms and checklists that could be employed during OT&E.

So, as part of the effort under Essex's contract to improve OT&E methodology, Essex Corporation was asked to evaluate the Quick HRTES to determine if the handbook would meet the needs of test officers conducting human factors tests and evaluations in a variety of settings. The evaluation also was to consider whether or not the HRTES volumes should be further revised. If revisions were required, the evaluation was to indicate what revisions should be made.

To accomplish the HRTES evaluation, a questionnaire was designed and administered to obtain the opinions and comments of experienced human factors professionals. A panel of twenty (20) experts initially agreed to participate by completing the questionnaire. There were fifty-four (54) closed-end items. The questionnaire used a 5-point rating scale, with anchors ranging from "very adequate" to "very inadequate." Seventeen (17) respondents replied. Fourteen (14) respondents completed the questionnaire, and three (3) other respondents provided extensive written critiques of HRTES.

Although disagreement among the reviewers was expected, the extent of the disagreement was not anticipated. On 28 of the 54 multiple-choice items, at least one respondent chose each of the five alternatives. This suggested that panel members may have had different expectations for HRTES, different views of the capabilities of test officers in the human factors area, or different views of the resources (especially time) available to the test officer in a typical operational test. However, overall, intra-rater agreement was fairly substantial ($r=.79$).

The comparatively low overall mean rating of adequacy ($m=3.32$ on a 1 to 5 scale, with 5 equal to "very adequate") indicated general disappointment with the Quick HRTES on the part of a large proportion of the reviewers. The highest mean rating for any chapter was 3.72, which still fell between "adequate" and "in between" on the rating scale. These ratings made it obvious that, in the opinion of many of the panel members, substantial revisions would be required to make HRTES acceptable.

Despite disagreements in ratings, there was considerable consistency among the comments received. In brief, four major concerns were expressed by substantial proportions of the panel: (1) terms were not well defined throughout the documents, (2) examples of recommended procedures were nonexistent or not adequate, (3) many procedures were too time consuming for a Test or Evaluation Officer and would produce results of doubtful utility, and (4) procedures, concepts, and criteria were not well tied together. For example, the reviewers felt that guidance was poor on what specific human factors measures should be selected in seeking to identify causes of performance failures.

Although the review panel did not give HRTES good marks, several panelists mentioned that the job it attempted to do bordered on the impossible. More detail might have made HRTES more acceptable to the panel; however, the prior, more detailed version of HRTES that was available had been rejected. Any attempts to revise HRTES following the review panel's suggestions would have resulted in an increase in HRTES' size. Several panelists, despite giving a number of very low ratings, desired to have a revised HRTES made available. They felt that a few revisions addressing some major shortcomings would make HRTES at least minimally acceptable.

The written evaluation submitted by Essex Corporation provided a more detailed account of the panel's views. The question of how to provide test officers the information and material needed to improve human factors evaluations of systems and equipment during OT&E has yet to be resolved. Subsequent Task 3 activities were directed at developing some answers.

Questionnaire Construction Manual

Under a previous contract with ARI, another contractor, Operation Research Associates (ORA), had reviewed the literature on questionnaire and interview construction, and survey administration research. In 1975, they produced two products: a Questionnaire Construction Manual (P-77-1) and a Questionnaire Construction Manual Annex: Literature Survey and Bibliography (P-77-2). The Questionnaire Construction Manual was revised/edited by Charles O. Nystrom, Ph.D., Fort Hood Field Unit, Army Research Institute (ARI), in 1976. In keeping with the Task 3 focus on improving OT&E methodology, Essex Corporation in 1983 was tasked to update these two publications. A search of the literature was initiated for questionnaire research done starting at ORA's cut-off date of 1973 and continuing to 1983.

The first Essex-produced volume was a revised Questionnaire Construction Manual; it has the same purpose as the manual it replaces. It is primarily for the guidance of those who are tasked to develop and/or administer questionnaires as part of Army operational tests and evaluations, such as those conducted by the TRADOC Combined Arms Test Activity (TCATA) and the Operational Test and Evaluation Agency (OTEA). The general content and concepts, however, are applicable to more than operational test situations. Thus, the manual should prove useful to all individuals involved in the construction and administration of surveys, interviews, or questionnaires. It provides guidance for performing each of the many steps that are involved in such undertakings.

The content of the revised Questionnaire Construction Manual covers questionnaire types, administration procedures, the development of questionnaire items, types of questionnaire items, content of questionnaire items, attitude scales and scaling techniques, response anchoring and response alternatives, format considerations, pretesting, interviewing, demographic characteristics, and evaluation of results.

The second volume produced was titled, Questionnaires: Literature Survey and Bibliography, is a sequel to P-77-2, not a revision. The volume is directed toward those who are tasked with questionnaire construction research ranging from research design and scale development through consideration of demographic characteristics of respondents. It is a summary and discussion of the relevant literature that appeared between 1973 and 1983.

To begin the update, a computer-assisted literature search was combined with a manual search and 16,816 citations were obtained; of these, 343 citations were identified as being potentially appropriate for questionnaire research. Subsequently, 178 citations were used as sources in writing the sequel, although 463 citations on questionnaire methodology are included in the bibliography.

Material for the sequel was researched and written using the actual journal articles, reports, and books, and not the abstracts of the journal articles. Journal articles, reports, and books selected for inclusion in the bibliography were screened for their relevance to questionnaire construction. The sequel was designed to answer questions about the latest technical methods for developing questionnaires. Questionnaires are used to assist Army personnel in performing field test evaluations. Methodological considerations which are relevant to constructing questionnaires, and which could be generalized from other fields for military application, were used in conjunction with questionnaire construction research from the military. Relevant literature was included for questionnaire construction research from other fields: political science, marketing, organizational management, human factors engineering, psychology, and education. Research on questionnaires was compared according to description of subjects, number of subjects, number and type of experimental conditions, number of scale dimensions, number of scale points, response alternatives, hypotheses tested, results, scale reliability, and scale validity.

Each section in the literature survey sequel was divided into four parts: (1) description of the content area, (2) examples of the content area, (3) comparison of studies, and (4) conclusions generated from the technical review. There were 27 different sections. Each section may be considered a stand-alone section, since it is possible to refer to an individual section for guidance in that content area without having to read the other sections and other chapters.

The literature survey volume was organized into several topic areas: (1) scale categories, (2) behavioral scales, (3) questionnaire item design, (4) scale design, (5) interview and respondent characteristics, (6) questionnaire format, and (7) suggestions for future research. There are detailed discussions within each topic area. For example, the discussion of scale categories contains an overview for various multiple-choice scales that represent nominal, ordinal, and interval measurement. The assumptions underlying scale construction and developmental procedures were reviewed for bipolar, semantic differential, rank order, paired-comparison, continuous, and circular scales. The section on behavioral scales consisted of a wide variety of forms and methods to develop scales which have behavioral anchors. The developmental procedures for behavioral scales were addressed. The discussion of questionnaire items design expanded upon contingencies involved in developing questionnaire items, such as the effectiveness of using positively and negatively worded items to create a balanced survey instrument. Other considerations included the number of items to use in a survey, and how many words to include in a question stem. The discussion of design of scale categories presented information on the selection of number of scale points and type of response alternatives.

Content covered regarding interviewer and respondent characteristics viewed questionnaire construction from the standpoint of the impact on the target population, as well as on the interviewer, instead of the impact ofthe design of the instrument. Demographic characteristics which influence item responses were examined. The section on questionnaire format focused on the physical structure of the questionnaire, the actual layout of the format, and the use of branching.

TCATA Test Officer's Training Manual

A TCATA Test Officer's Training Manual had been distributed in November, 1979. A review and revision of the manual (called Appendix G) was requested by ARI in 1983 as part of Task 3. The purpose of the review was to determine whether or not the 1979 edition retained its theoretical soundness regarding interview and questionnaire methodology in light of the research available in 1983. The objective of the project was to revise Appendix G where applicable to reflect the state-of-the-art in questionnaire methodology.

Overall, Appendix G was found to be theoretically sound. Only minor changes were recommended. Original topic areas found in the 1979 text were supplemented, and no major revisions were suggested. Review of the literature, and resulting recommendations, along with some specific revisions were provided. Revisions to Appendix G, Interview and Questionnaire Methodology covered topic areas for branching, threatening or sensitive questions, "don't know" response sets, awareness level of respondent, confidentiality, number of scale points, scaling considerations, and anchors.

Training on OT&E Interview Techniques

Interviewing is a common technique for the administration of surveys to measure individual and system performance during operational testing. Such interviews can be biased in a number of ways. For example, how the interviewer orally identifies the scale points and how the questions are presented to the respondent have the potential to bias survey research results. Interviewers may inadvertently suggest answers to those being interviewed. These are a few of the ways in which questionnaire research and the information that results from it can be influenced during the interview process. In other words, the behavior of untrained interviewers may detract from the standardized administration of questionnaires.

It was determined that providing training for OT&E interviewers could improve standardization of field administration of questionnaires and control for methodological bias. Indeed, interviewer training was considered essential for maintaining content validity and reducing potential interviewer bias during OT&E interviewing.

A videotape presentation was selected as a potentially useful and effective way ofproviding such training and thereby reducing errors in questionnaire administration. Therefore, in 1985, a 30-minute video was produced to train civilian and military personnel who conduct operational test and evaluation interviews for the Army Research Institute and TRADOC Combined Arms Test Activity (TCATA). The objective of the video training was to train military and civilian personnel in how to conduct a standardized OT&E interview, and toincrease confidence in operational test and evaluation findings. The overall purpose was to improve interviewing skills for U.S. Army civilian and military personnel.

In preparing the video, the research team responsible for producing it formulated an instructional approach along with an analysis of the training needs of civilian and military personnel. The population was defined in terms of interviewing experience, educational background, attitudes toward training, desired types of improvement in interviewer competence, etc. Course objectives were developed for operational test and evaluation interview training. Content of the video training included topic areas such as situational factors found in interviewing, standardized interviews, introduction of a questionnaire, probing respondents, and recording answers. The video demonstrated how operational test and evaluation interviews should be conducted in the field and in an office setting. The video was shot on location at Fort Hood. Field locations showed actual Army equipment and systems, i.e., Bradley Fighting Vehicle, M1 Abrams Main Battle Tank.

Seven sequences were incorporated into the video. The first sequence introduced the importance of the interview in meeting the goal of the OT&E program. The second sequence demonstrated how to prepare for the interview, followed by a third sequence which described the important initiating steps in conducting an interview. The remaining four steps portrayed actual interviews being conducted. The scenes featured interviews primarily in the field situation with "local talent" demonstrating such skills as probing for more complete answers, reading items slowly, recording responses, and concluding aninterview. ARI furnished all necessary on-screen talent; no professional actors were employed.

Operator and Maintainer Workload Assessment

Assessing workload, knowing how much work a system requires of an operator or maintainer, or knowing how much work an operator or maintainer can do, is an important part of operational testing. For more than a decade, workload, and particularly mental workload, has been an extremely busy research area. There was a great quantity of research and a number of reviews of that research. Despite the research activity, mental workload remains more a credible concept than a real-world applicable measure.

Assessing workload is an integral part of the operational test and evaluation situation. Degradation in individual performance, and in system performance as a consequence, may be due to operator overload. Equipment or system unavailability may be due to maintainer overload. If operators or maintainers are underloaded, personnel costs are higher than they need be. Thus, in the interests of system performance and cost effectiveness, workload assessments need to be as accurate as possible.

Despite the need for workload measures, there is generally agreement that there is not even an accepted definition of the term "mental workload." It has been equated with the time taken to perform a task, the arousal level of the operator, or the subjective experience of cognitive effort. Attempts to measure workload include measurements of primary task performance, secondary and dual task measures, subjective measures, and physiological measures such as heart rate, pupillary response, and evoked brain potential.

For those seeking to apply workload assessment in the context of OT&E, the volume of the available research presented a problem as well as an opportunity. Automated searches produced many irrelevant citations. In part because of widespread concern with workload, workload often was used as a key term to note the inclusion of speculative discussion on the topic, as well as data on it. Thus, anyone seeking references on workload measurement that would be useful in an OT&E context could easily be inundated with material and yet be short of information.

As a Task 3 project aimed at improving the methodology available for human factors aspects of OT&E, it was decided to create a special purpose bibliography on workload assessment. A literature search was initiated with the narrowed focus of workload assessment for operators and maintainers in the context of OT&E. Although some automated searches were conducted, the major effort was directed toward manual searches, since some of the most fertile sources of potentially useful papers were not included in the automated databases. The emphasis in the search was on papers published during the last decade.

The objective of this project was to provide a starting place for those seeking to develop better methods of workload assessment for application in OT&E environments. From the hundreds of papers on workload that were reviewed, 301 were selected for their relevance to some aspect of operator or maintainer workload in an OT&E setting. Abstracts of these papers were arranged alphabetically by author and compiled into a loose-leaf volume. In addition to the abstracts, this volume included the tables of contents of four relevant books. To increase the availability of useful information, 81 of the papers judged to be most useful were collected into a second volume.

The form of these two volumes (Seven, 1986) recognizes the variety of needs and purposes of those who deal with the concrete problems of workload measurement. They represent a distillation of a massive amount of material. The workload assessment volumes were designed to be working tools. Every time someone consults either volume, the specific current interest and aim that motivates the search will also inform and color what they get out of the search.

Workload measurement will continue to challenge those who deal with system operators and maintainers for years to come. It is our hope that these two volumes will make the search for measures applicable to real-world, real-time systems more effective.

Human Factors Analysis of the DIVAD System

This project resulted in two volumes presenting a Human Factors analysis of the Division Air Defense (DIVAD) Gun System, also known as Sgt York. The first volume was a consolidation and analysis of the human factors data obtained from the Sgt York Follow On Evaluation I (FOE I) tests. The second volume was a discussion of the lessons learned from that FOE.

Although MANPRINT requirements were not imposed on Sgt York during FOE I, the six domains were used as an organizing basis for examining the outcome of FOE I in Volumes I and II. The six MANPRINT domains are: (1) human factors engineering, (2) manpower, (3) personnel, (4) training, (5) system safety, and (6) health hazards. Issues identified within the six categories were addressed in both volumes.

The purpose of Volume I (Babbitt, 1987) was to document the Follow-On Evaluation tests conducted between April 1985 and June 1985 to support an assessment of the DIVAD Gun System, the Sgt York. The Force-on-Force phase of FOE I was conducted at the Combat Development Experimentation Center (CDEC) at Fort Hunter-Liggett, CA, and the Live Fire phase was conducted at the White Sands Missile Range in New Mexico.

Essex Corporation was under contract (MDA903-85-C-0229) to the U.S. Army Research Institute for the Behavioral and Social Sciences to carry out human factors, training, and safety analyses of the Sgt York. Mr. George Gividen, Chief of the ARI Field Unit at Fort Hood and ARI coordinator for human factors on the Sgt York FOE I test, was the Contracting Office Technical Representative (COTR) for that contract. A seven-man Essex human factors team was on-site as the Force-on-Force and Live Fire phases of the Sgt York FOE I tests were conducted. A preliminary account of the human factors, safety, and training results of FOE I was supplied for incorporation in the Operational Test and Evaluation Agency (OTEA) report on FOE I. Those results also provided the foundation for the two-volume work. Actual report preparation was covered under Task 3 since the focus was on improving OT&E methodology.

The objective of the first volume was to consider human factors, training, and safety deficiencies found to exist during the FOE. Human factors, safety, and training problems noted during FOE I were identified and clustered into twelve subcategories: (1) Physical Environment and Workspace; (2) Workspace, Anthropometrics, Comfort; (3) Controls and Displays; (4) Workload/Division of Labor; (5) Visibility; (6) Audio and Visual Alarms; (7) Target Detection/Acquisition/Tracking; (8) Communications; (9) Travel/Navigation; (10) Publication/Documentation; (11) Safety; and (12) Training. When the seriousness of the impact was rated for the problem by subcategory, the average impact across all 12 categories would have predicted less than optimal mission performance. Average ratings by subcategory indicated that there were four areas which were considered as seriously degrading mission performance. The subcategories identified were Physical Environment and Workspace, Workload/Division of Labor, Target Detection/Acquisition/Tracking, and Travel/Navigation. Findings indicated that redesign of the system, its components, and crew workspace would have improved the work environment, enhanced the system, and removed deterrents to good system performance.

The purpose of the second volume, Lessons Relearned (Seven, 1987), was to examine lessons that could be learned from the human factors aspects of the Sgt York Follow-On Evaluation (FOE I) tests. The report focused on (1)applying the experience of Sgt York FOE I to the larger issue of integrating good human factors design into the entire process of weapon system acquisition: (2) using the Sgt York FOE I observations to suggest improvements in human factors operational test and evaluation; (3) putting these observationsinto the context of the current Army wide MANPRINT initiative; and (4) relating findings to the results of earlier reverse engineering and design criteria studies of other major Army weapon system acquisitions (Stinger, Multiple-Launch Rocket System (MLRS), Fire Support Team Vehicle (FIST-V), and the Fault Detection and Isolation Subsystems of the M1 Tank.

The report provided brief descriptions of Forward Area Air Defense (FAAD), of the intended role of the Division Air Defense (DIVAD) Gun in the FAAD mission, of the acquisition history of DIVAD, and of the resulting DIVAD gun system and its operation. The human factors requirements that were imposed on the Sgt York system and the tests conducted on that system were reviewed. Encompassed within the broader human factors domain were issues of training, safety, manpower, and personnel, as well as human engineering concerns associated with displays, controls, workspace, and the like. Not all these issues were addressed in the requirements documents, but specifications associated with any of these human factors considerations were noted in the report.

Lessons to be learned from the DIVAD program were considered in conjunction with the results of earlier reverse engineering studies and of other human factors operational tests that were applicable to future FAADS developments. The report (Seven, 1987) addressed issues of continuing relevance to the system acquisition process and to the successful integration into that process of the concerns which underlie the MANPRINT initiative.

The report concluded that adequate design criteria need to be available; once available, they need to be used. Ensuring and facilitating the use of design criteria appeared to be the most urgent need. Clearly the fact that the pertinent criteria were available in a Military Standard (MIL-STD) or Military Handbook (MIL-HDBK) was not enough. Somehow, such criteria must become meaningful and familiar to weapon system design teams, if not to everyone on such a team, at least to some. Failure to meet human factors criteria must be acknowledged for the serious problem it is. Making adequate human factors criteria available is only the first step in a progression that must include applying them; that was one of the lessons that should not have had to be relearned. It is a lesson that will continue to be repeated until it is learned.

Findings from the report indicated that FOE I preparations were rushed and incomplete; training was abbreviated; data collection and analysis were hurried, incomplete, and ad hoc. Too much had to be done too quickly; much was done but much was left undone. Sgt York crewmembers were asked to operate a new, complex weapon system with too little training and too much stress. Test scenarios did not reflect adequately the problems suggested by earlier tests or the strengths built into the weapon system, partially if not primarily because of lack of familiarity with either one. Many unanswered questions remained which were substantial in nature (Seven, 1987).

Principal Products

The six Task 3 projects identified and described in this chapter have led to nine (9) products (eight reports and one video tape. All nine products listed below were prepared for ARI by Essex Corporation at its Fort Hood, TX facility or at the Westlake Village, CA office.

Babbitt, B. A. (1983, September). Revision of the TCATA test officer's training manual, Appendix G.

Babbitt, B. A. (1987, June). MANPRINT analysis of the DIVAD system: I. Human factors data from Sgt York follow on evaluation I.

Babbitt, B. A., & Nystrom, C. O. (1985, January). Questionnaires: Literature survey and bibliography.

Babbitt, B. A., & Nystrom, C. O. (1985, March). Questionnaire construction manual.

Babbitt, B. A., Semple, C. A., Sparks, R. J., Seven, S. A., Sharkey, M., Shaw, B., & Wright, C. L. (1986, June). Operational test and evaluation interviewing techniques video training (Video).

KubaJa, A. L. (1985, December). Evaluation of human resources guide book for operational test officers.

Seven, S. A. (1986, May). Operator/maintainer workload assessment: Volume I. Selected literature sources.

Seven, S. A. (1986, May). Operator/maintainer workload assessment: Volume II. Selected articles.

Seven, S. A. (1987, June). MANPRINT analysis of the DIVAD system: II. Lessons relearned.

Continuing Methodological Questions

Review of the six research projects carried out under Task 3 and discussed above indicated that there remain unanswered methodological questions. As a result of these projects aimed at improving OT&E methodology, some methodological areas revealed opportunities for profitably pursuing certain lines of research. Recommendations are provided here for following up on these opportunities. Where the potential for payoff is limited, recommendations are either minimal or are not presented. Suggestions on future research to answer methodological questions are presented for the Human Resources Test and Evaluation System (HRTES), for content areas associated with questionnaire development, for operator/maintainer workload assessment, and for MANPRINT implementation of operational test and evaluation.

Human Resources Test and Evaluation System (HRTES)

If HRTES were to be revised, procedures already identified for simplifying the document or making it easier to use should be followed. Resources typically available to test officers were more limited than the intricate and lengthy HRTES procedures. However, since the implementation of MANPRINT methodology for operational test and evaluation, OT&E procedures have actually become even more complicated. See Chapter VI, Follow On Evaluation Procedures, of Babbitt, Seven, and Muckler (1987) for further discussion.

HRTES consists of two volumes. Perhaps the supplemental part (second volume) could be integrated into the text of the first volume so that the reader need not search back and forth between the two volumes. The content of HRTES purported to cover solutions to problems in the areas of tactics, doctrine of employment, and organizational structure. The content area was dealt with on such a limited basis that it did not provide enough information to be useful. A revision would be enhanced by eliminating discussions of tactics, doctrine of employment, and organizational structure. Those topics are not directly relevant to conducting human factors evaluations of military equipment. Expanding discussions of such tangential content areas so that test officers could actually use the guidance to solve problems would make the handbook more cumbersome and would not help it to better meet its original purpose.

Specific suggestions to improve the quality of HRTES were incorporated into the report "An Evaluation of a Human Resources Guide Book for Operational Test Officers." For example, the procedures volume indicates that human factors experts need to take some measurements when conducting human factors test and evaluation, but more detail should be included in the supplement section on matters such as which measures should be used. The report also suggested that the HRTES volume should state how and under what conditions human factors expertise should be obtained. The report contained further suggestions for a chapter-by-chapter revision of HRTES.

Questionnaire Development

The 1983 search of the literature to update information on how to construct, administer, and analyze questionnaires indicated that questionnaire construction research had progressed unevenly across professional disciplines. Methodological considerations for questionnaire construction require a comprehensive series of experiments, yet there has been a paucity of sustained research in this area.

A new trend in questionnaire research has emerged in the past few years. Computers have transformed questionnaire construction, administration, and scoring. Efficiency and economic advantages have resulted from the application of computers to questionnaire research.

In the literature survey volume on questionnaires, future research topic areas were identified for further investigation. These topic areas included scale development procedures and analysis, procedural guides to item wording, subjective workload assessment methods, Behaviorally Anchored Rating Scales (BARS), item non response, branching, demographic characteristics, and pictorial anchors.

Research recommendations were selected for their relevance and application to OT&E activities. Within the last few years there has been a shift in research focus. Previous questionnaire research was concerned with variables such as continuous scales and discrete scales, response alternatives, number of scale points, type of scale format, etc. Conflicting research results seem to indicate that each different scale format has its own strengths and weaknesses. More recent investigations of other variables have focused on survey developmental procedures, adaptive testing formulated as a computer survey, expert systems, and characteristics of respondents, including their cognitive complexity. Some suggestions for future questionnaire research are presented below.

Scale Development Procedures and Analyses

Military survey research for the OT&E community needs to investigate ways to obtain more lead time in survey development. Item reduction and multidimensional scaling techniques have been used in commercial-industrial surveys which may be applicable for Army surveys. This could be a vehicle to introduce scale development procedures that would reduce the number of items used in field surveys. In conjunction with scale development procedures, statistical analyses should compare different formulas and statistical assumptions.

Procedural Guides to Item Wording and Expert Systems

There is no consensus among survey researchers as to how to word items, or the tone of the wording. Procedures have been developed to help identify what specific words should and should not be used in an item. Various procedures for identifying the specific words to be used in an item could be compared; some procedures possibly may identify the structure of the item itself. A method for selecting item wording has been developed to ensure that respondents only be subjected to items they can understand. It may be possible to incorporate the procedures and decision-making processes of this method into an expert system using higher-order languages. Items generated by an expert system would still require extensive pretesting.

Automated Portable Test Systems

Administration of surveys could be conducted on a portable test system using a microprocessor which is user-friendly. Entering and collating responses need to be performed with accuracy and precision. Portable systems can be used simultaneously at various remote sites and incorporated into a real-time network. Questionnaires can be constructed for (and eventually by) this type of automated system. Preliminary development of such systems already has been done.

Item Non-Response, Branching, and Demographic Characteristics

Branching offers considerable potential for survey efficiency. The relationships between item non-response, branching, and demographic characteristics for a military sample are among topics that should be considered in developing new questionnaire formats.

Test Officer's Training Manuals

Appendix G, the TCATA Test Officer's Training Manual, was distributed in November, 1979, as previously noted. The recommendations submitted in 1983 for revision of this manual reflected minimal change. However, future revision of Appendix G might address the issues of reliability, validity, and special problems associated with different measurement scales. Such a revision would require the inclusion of new material into the text. Appendix G did not cover these issues since not all test officers have had course work in measurement. Topic areas related to psychometrics and specific types of data analysis could be incorporated into the appendix if appropriate and specific procedures are provided as guidelines for when and how to select and apply measures.

The TCATA Test Officer's Training Manual, Appendix G, and the Questionnaire Construction Manual both could be revised to promote a better technical understanding of questionnaire methodology for the portion of the audience who are uninitiated in this field. It is understood that specific guidance regarding the interpretation of the analysis of questionnaire data may be desirable and could enhance the viability of the manual.

Questionnaires are useful in obtaining information, but combining supplemental measures of performance with questionnaires is preferable. Direct observation is the most natural method of gathering information, and observational techniques are frequently combined with questionnaires. Combinations of data-gathering techniques for field testing have included establishing critical task crew performance measures, critical task assessment interviews, on-site observations, structured interviews covering test data requirements, and collecting comment and opinion data from test participants, as well as using other techniques such as video and audio recordings.

Illustrations to provide guidance in questionnaire construction areas already have been incorporated into the manual. However, there are further illustrations that could be generated which would enhance the manual's effectiveness. In addition, methods which have traditionally been used regarding questionnaires and testing of Army equipment have been changing under the influence of the MANPRINT initiative.

If test officers' training manuals are to provide adequate guidance to individuals who are tasked with developing questionnaires, they would benefit from the addition of a chapter on the topic of the role of questionnaires and the MANPRINT process which involves human factors, safety, health hazards, manpower, personnel, and training considerations. Content of the chapter should focus on crew performance data collection that is preplanned to accommodate the system as it is designed and built. Crew performance data collection considerations could emphasize preplanning activities conducted prior to testing. Discussion of measurement technology to accommodate the measurement of both crew and system performance are viewed as an integral part of the process. The use of combined data collection techniques such as timed crew performance measures on critical tasks, and the application of stringent criteria to critical crew performance tasks, could be included. Other topics covered should pertain to issues such as the training of human factors observers, how observations are taken and recorded, what instruments are needed and available, and how the data may be recorded, analyzed, and reported.

Video Training and Interviewing

The video training produced to train OT&E interviewers to standardize field administration and to control for methodological bias associated with interviewing was one step along the way to transforming traditional methods of instruction into more visual approaches for the trainee. Since some of the content may be complicated for individuals who lack a scientific background, a multimedia instruction approach could be quite useful. For example,video presentations could be supported by instructor guides. Using this media mix would provide greater interaction between trainers and trainees. It would allow for greater interaction and feedback. Prior to the production of subsequent videos, a needs analysis should be conducted for any proposed questionnaire content areas, and levels of performance for competencies should be identified. Course objectives need to specify the instructional outcomes.

Operator/Maintainer Workload Assessment

There are continuing methodological questions associated with both the definition and the assessment of workload, and these questions are at least as pressing in the context of operational test and environment as they are in other arenas. Just what constitutes workload seems to depend both on what is being assessed and on who is doing the assessment. Fur thermore, workload is a function not only of the tasks to be done and of the environment in which they are to be accomplished, but also of the skills, training, and experience of the doer. Workload measures need to be developed that are sensitive to all of these variables.

Future methodological investigations are needed in support of OT&E in at least three specific areas:

- Transferring workload assessment methods from the research laboratory to the operational context appropriate to the test and evaluation environment.

- Adjusting for the sensitivity to individual differences that exists in current workload estimation techniques.

- Developing workload assessment based on real-time individual performance in context.

A number of specific investigations should be undertaken in beginning to implement these investigations. For example, the concepts of primary and secondary task measures should be modified to fit the operational context of OT&E. In addition to primary and secondary task measures, there are another two major types of workload measures, physiological and subjective measures. The applicability of these two measures should also be considered.

Just as physiological and subjective workload assessment techniques are worth pursuing because they appear to offer composite estimates or measures which integrate the effects of multiple demands, so too some empirical assessments based on real-time performance measures should be investigated. In many automated and semi-automated systems, system performance measures are available that reflect the performance of the on-duty officer. Such measures would reflect individual differences in performance and provide another approach to workload assessment.

Analysis of the DIVAD System

The questions about Sgt York that remained unanswered were substantial. Some findings from Volume I led to the following general conclusions: Human factors, safety, and training design criteria were inadequately imposed on the design of the Sgt York Air Defense System. The weapon acquisition process was accelerated, negatively influencing the resolution of human factors, safety, and training problems identified in previous DT/OT evaluations. Training efficiency and effectiveness for the FOE I tests were negatively impacted by the acceleration. Test trials were constrained to intervals of 20-30 minutes due to the instrumentation used in data collection. As a result, the significance of human performance problems was underestimated. If operation had been sustained for 72 hours during FOE I as originally projected, human factors, safety, and training problems seriously degrading combat system performance would have been certain.
While the proportion of avoidable problems varies from system to system, it seems clear that there is at least as much problem with criteria not being applied as with their not existing. In far too many instances, had existing criteria been followed, the resulting systems would have had far fewer problems. Systems are still being presented for operational test and evaluation that grossly violate existing human engineering standards (MIL-HDBK-759A, MIL-STD-1472C). Performance degradation is predictable and, during OT&E, it is apparent. For example, when crew stations are too small to accommodate crewmembers or their necessary actions, system performance suffers, to say nothing of the operators.

If the goal of OT&E were only to find and document deficiencies in system and crew performance, there would be no need to be concerned with methodological issues such as how to assure compliance with human engineering standards or how to increase the prerequisite familiarity with such standards. If the goal of OT&E is to find deficiencies, the fact that the standards are often not followed assures that there will be plenty of deficiencies and problems to find. In the hope that OT&E is intended to improve systems as well as to warn users of their shortcomings, we urge that the continuing methodological problem of implementing existing human engineering design criteria be recognized and addressed.

In the early 1980s, several case studies of the development of specific weapon systems were conducted. These studies led to the identification of explicit deficiencies in considering man as a system component, and in integrating human performance into the total operational performance of a fielded situation (Seven, 1987). Through reverse engineering studies and a series of reports produced by ARI entitled "Human Factors Engineering Design Criteria for Future Systems," emerging methodology has been developed (Crumley& Earl, 1985; Daws, Keesee, Marcus, Hartel, & Arabian, 1984; Earl, 1984; Earl & Crumley, 1985; Promisel, Hartel, Kaplan, Marcus, & Whittenburg, 1985). The adequacy of existing design criteria has been examined as part of this process. There has been a continued focus on the relationship of design criteria to problems encountered in various component systems during field testing.

Reverse engineering studies repeatedly document problems that were easier to recognize than to fix. In attempting to take the step from recognition to resolution of these problems, General Maxwell Thurman (then Vice Chief of Staff, U.S. Army) sponsored the MANPRINT initiative. The concerns are not new, and MANPRINT is not the first attempt to improve the systems that are developed by looking early, in depth, and in context at the impact of human factors on systems. Prior efforts to integrate the consideration of such factors into the weapon system development process can be traced back more than two decades and can be identified in each of the services; HARDMAN was the most recent predecessor. Each successive effort has attempted to increase the breadth and sophistication of awareness of the need to integrate human factors into systems design and development. MANPRINT is the latest of these efforts and it appears to be the most insistent.

Perhaps as demands for compliance multiply and as implementation becomes more concrete and customary, MANPRINT efforts will begin to improve the systems that are fielded and the operational tests that are conducted to evaluate these systems. Nonetheless, embedding human factors awareness into all phases, and especially into the early phases, of weapon system development will continue to pose challenges to system designers and human factors specialists alike.

Program Personnel

Individuals responsible for producing products under Contract No. MDA903-83-C-0033, Task 3: Improving Human Factors (MANPRINT) Test and Evaluation Methodology, are listed below. All identified individuals are Essex Corporation employees with the exception of one consultant (Mr. Sharkey).

Dr. Bettina A. Babbitt
Dr. Albert L. Kubala
Dr. Frederick A. Muckler
Mr. Clarence A. Semple
Dr. Sally A. Seven
Mr. Michael Sharkey
Mr. Brian E. Shaw
Dr. Ronald J. Sparks
Mr. Charles L. Wright