NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

AC/323(HFM-101)TP/52

www.rta.nato.int

RTO MEETING PROCEEDINGS **MP-HFM-101**

Advanced Technologies for Military Training

(Technologies avancées pour l'entraînement militaire)

Papers prepared for the RTO Human Factors and Medicine Panel (HFM) Symposium held in Genoa, Italy, 13-15 October 2003.

Published April 2004

Distribution and Availability on Back Cover

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

AC/323(HFM-101)TP/52

www.rta.nato.int

RTO MEETING PROCEEDINGS **MP-HFM-101**

Advanced Technologies for Military Training

(Technologies avancées pour l'entraînement militaire)

Papers prepared for the RTO Human Factors and Medicine Panel (HFM)
Symposium held in Genoa, Italy, 13-15 October 2003.

The Research and Technology Organisation (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote co-operative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective co-ordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also co-ordinates RTO's co-operation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of co-operation.

The total spectrum of R&T activities is covered by the following 7 bodies:

- AVT Applied Vehicle Technology Panel
- HFM Human Factors and Medicine Panel
- IST Information Systems Technology Panel
- NMSG NATO Modelling and Simulation Group
- SAS Studies, Analysis and Simulation Panel
- SCI Systems Concepts and Integration Panel
- SET Sensors and Electronics Technology Panel

These bodies are made up of national representatives as well as generally recognised 'world class' scientists. They also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier co-operation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

The content of this publication has been reproduced
directly from material supplied by RTO or the authors.

Published April 2004

ISBN 92-837-1117-3

Advanced Technologies for Military Training

(RTO-MP-HFM-101)

Executive Summary

From 1985 through 1995, NATO held a series of workshops on technologies associated with military training, covering Computer-Based Instruction, Training System Design, Distributed Training, and Virtual Reality. While technologies have dramatically changed, many of the problems and issues related to training technologies, their application, acceptance, and effectiveness, are still with us today. The re-structuring of military forces in size and composition, along with decreased funds for real exercises, has meant that attention has remained focussed on ways of maintaining a well-trained and flexible force. Recent changes in NATO and in the European infrastructure stimulated the HFM Panel of the RTO to sponsor a symposium focused on technologies applicable to the improvement of military training.

The symposium illustrated four themes. The first one was introduced by Dr. Dee Andrews, co-chairman of the symposium, and Admiral Accardo. They set the tone, the theme, and the challenge for the NATO R&D community. The challenge illustrated by Dr. Andrews' talk is the sharing of information among alliance and partner nations in NATO. The suggestion made in his talk, implicitly if not explicitly, is that this multinational organization should surpass its present cultural, economic, business, and internal borders, and develop a supra national interdependent identity outside the military domain. Such an organization would be open and in the context of Andrews' speech, a learning organization. With the availability of advanced technologies and the opportunity for creating truly multinational synthetic training environments, NATO should now examine how to accomplish cross-cultural sharing of information, facilities, and training goals. Admiral Accardo's video, demonstrated that at least among seven nations, an effective joint training exercise could be achieved by the sharing of mission training rehearsal information and the creation of a combined training environment, with advanced technologies, using an embedded training approach.

The second theme was illustrated by two presentations dealing with the Project "First Wave", which illustrated the exciting potential to actualize the promise intrinsic to Dr. Andrews theoretical proposition. These presentations accomplished two things. First, they clearly identified the technical and technological obstacles that must be overcome, when asking several nations to share mission training exercises with simulators and various types of synthetic opposing forces (the interoperability issues). They also showed that these technological obstacles even in a preliminary attempt at orchestrating this mission training can be overcome by goodwill and cooperation among nations. On the other hand, they also highlighted the need for nations to overcome their insecurities and cultural differences concerning the sharing of information. In fact, the principal difficulty seemed to be, establishing a way for each nation to both share information and facilities with its fellow nations in NATO, and at the same time retain an acceptable degree of security for itself.

The third theme was illustrated by the research presentation concerning the evaluation of the Virtual Environment (VE) for aircraft maintenance training. The value of this study was in its careful attention to experimental detail in conducting research to evaluate both the effectiveness and the cost benefit of using the virtual environment for training. The research team specifically asked the question whether or not VE would be useful in solving a real training problem. They set about conducting a front-end analysis, completed a systems approach to training, and evaluated their product in the real world. In terms of its attention to detail, careful consideration of experimental design, and concern for practical as well as statistical significance, this research project could serve as a model for other nations to follow.

Finally, although only one presentation concerned itself with experimental validation involving the teaching of critical thinking to leaders, a number of papers concerned the need to shift from procedural kinds of training to a much more holistic, cognitive development approach for training individual leaders and teams in decision making. Ironically, the field of military training, which usually leads the field of education in applying innovations in items like technology, could benefit from the years of educational research which have been spent on cognition and meta-cognitive development. In fact, one presentation specifically concerned the application of the theory of tacit knowledge, developed largely in educational environments, to military tactical decision making. The reason for the shift in emphasis to cognitive development seems to be the recognition that with the increasing complexity of using advanced technologies in cockpits, on shipboard, and in warfare in general, much more abstract, cognitive demands are placed on individual soldiers, leaders, and teams alike. The nature of likely conflict, which will probably be of high-intensity, frequent, and in unpredictable locations, combines to demand, flexible, self-sufficient, and highly mobile forces. Together, advanced technologies and the changing nature of warfare place increasing cognitive demands on all personnel.

Technologies avancées pour l'entraînement militaire
(RTO-MP-HFM-101)

Synthèse

De 1985 à 1995, l'OTAN a organisé une série d'ateliers sur les technologies associées à l'entraînement militaire, couvrant la formation automatisée, la conception des systèmes d'instruction, l'instruction à distance et la réalité virtuelle. Si, depuis, les technologies ont changé de façon radicale, bon nombre des problèmes liés aux technologies de l'entraînement, à leur mise en application, à leur acceptation et à leur efficacité persistent aujourd'hui. La restructuration des armées, du point de vue tant de leur taille que de leur composition, associée à la diminution des crédits disponibles pour les exercices en vraie grandeur, explique l'attention qui continue d'être portée au maintien d'une force armée bien entraînée et adaptable.Les changements intervenus récemment dans l'infrastructure européenne et à l'OTAN avaient incité la commission HFM de la RTO à organiser un symposium sur les technologies applicables à l'amélioration de l'entraînement militaire.

Ce symposium a abordé quatre thèmes. Le premier était présenté par l'amiral Accardo, ainsi que par M. Dee Andrews, coprésident du symposium. Ces deux orateurs ont donné le ton de la conférence, en présentant le sujet et le défi qu'il présente pour les chercheurs de l'OTAN. Il s'agit du partage de l'information entre les pays membres de l'Alliance et leurs pays partenaires. Au cours de sa présentation, M. Andrews a indiqué implicitement que cette organisation multinationale devrait dépasser ses limites culturelles, économiques, commerciales et internes actuelles pour construire une identité supranationale, indépendante, hors du domaine militaire. Une telle organisation serait ouverte et, dans le contexte de la communication de M. Andrews, elle serait également évolutive. Disposant de technologies avancées et ayant la possibilité de créer de véritables environnements d'entraînement synthétiques multinationaux, l'OTAN devrait désormais étudier la question du partage interculturel des informations, des installations et des objectifs d'entraînement. Le film vidéo projeté par l'amiral Accardo, a démontré qu'un exercice d'entraînement interarmées efficace pourrait être organisé entre sept pays au moins, moyennant le partage d'informations sur l'entraînement de préparation à la mission et la création d'un environnement d'entraînement multinational, incorporant des technologies avancées, selon une approche intégrée.

Le deuxième thème était illustré par deux présentations portant sur le projet « First Wave », qui a démontré les possibilités très intéressantes de concrétisation des activités évoquées par la proposition théorique du M. Andrews. Ces présentations ont atteint deux objectifs. En premier lieu, elles ont identifié très clairement les obstacles techniques et technologiques qui sont à surmonter pour obtenir la participation d'un certain nombre de pays à des exercices d'entraînement à la mission mettant en jeu des simulateurs et différents types de forces antagonistes synthétiques (questions d'interopérabilité). Elles ont également démontré, par le biais d'une première tentative d'organisation de l'entraînement à la mission, qu'il est possible de surmonter ces obstacles technologiques grâce à la bonne volonté et la coopération entre nations. Par contre, elles ont aussi souligné à quel point il est important pour les différents pays de maîtriser leurs appréhensions et leurs différences culturelles concernant le partage de l'information. En réalité, la principale difficulté serait de trouver le moyen pour chaque pays de partager de l'information et des installations avec les autres pays de l'OTAN, tout en assurant un degré acceptable de sécurité propre.

Le troisième thème était illustré par la présentation sur l'évaluation de l'environnement virtuel (VE) pour l'entraînement à la maintenance des aéronefs. L'intérêt de cette étude réside dans l'attention qui avait été portée au détail expérimental lors des travaux de recherche destinés à évaluer l'efficacité et la rentabilité de l'environnement virtuel pour l'entraînement. En l'espèce, l'équipe de recherche a posé la question précise de savoir si le VE pourrait être utile à la résolution d'un problème réel d'entraînement. Elle a réalisé une analyse préliminaire, formulé une approche système de l'entraînement et évalué son produit en situation réelle. Etant donné son souci du détail, son examen consciencieux de la conception expérimentale, sa recherche d'intérêt tant pratique que statistique, ce projet de recherche pourrait servir de modèle à d'autres pays souhaitant poursuivre sur cette voie.

Enfin, bien qu'une seule présentation ait été donnée sur la validation expérimentale comprenant l'enseignement de la pensée critique aux chefs d'équipe, un certain nombre de communications ont évoqué la nécessité d'accorder moins d'importance à l'entraînement procédural et davantage à un type d'entraînement plus holistique et cognitif pour la formation de leaders et d'équipes à la prise de décisions. Il est amusant de constater que le domaine de l'entraînement militaire, habituellement le premier à appliquer les innovations technologiques, pourrait bénéficier des années de recherche pédagogique consacrées à l'étude du développement de la cognition et de la métacognition. En effet, l'une des présentations portait spécifiquement sur l'application de la théorie de la connaissance tacite, développée en grande partie dans des établissements scolaires, à la prise de décisions tactiques militaires. Ce déplacement d'attention portée au développement cognitif s'expliquerait par la reconnaissance générale du fait qu'avec la complexité croissante des technologies avancées mises en œuvre dans les postes de pilotage, à bord des navires et sur le champ de bataille en général, le combattant individuel, les chefs militaires et les équipes doivent tous répondre à des exigences cognitives, de nature beaucoup plus abstraite. Les conflits futurs seront probablement fréquents, de haute intensité et se dérouleront dans des lieux imprévisibles. Il en résulte la nécessité de forces armées flexibles, autonomes et d'une grande mobilité. La combinaison des technologies avancées et l'évolution de la conduite de la guerre imposent de plus en plus d'exigences cognitives à l'ensemble des personnels.

Table of Contents

SESSION III: TEAM AND COLLECTIVE TRAINING

SESSION IV: TRAINING ANALYSIS AND EVALUATION

SESSION V: DISTRIBUTED SIMULATION

SESSION VI: COMMAND AND CONTROL/LEADERSHIP TRAINING

† Paper not available at the time of publishing

Human Factors and Medicine Panel

CHAIRMAN

Dr. Robert ANGUS
Director General
Defence R & D Canada – Suffield
PO Box 4000 – Station Main
Medicine Hat, Alberta T1A 8K6
CANADA

VICE-CHAIRMAN

Dr. Jean-Michel CLERE
Chef du Département Sciences Médicales et
Facteurs Humains DGA/DSP/STTC/SH
8, Boulevard Victor
F-00303 Armées
FRANCE

PROGRAMME COMMITTEE

CHAIRMAN

Dr. Dee ANDREWS
Senior Scientist
War Fighter Training Research Division,
United States Air Force Research Laboratory,
6030 S. Kent Street, Mesa, Az, 85212-6061
UNITED STATES

CO-CHAIRMAN

Mr. Rene LAROSE
Director General DRDC Toronto
1133 Sheppard Avenue West,
PO Box 2000
Toronto, Ontario M3M 3B9
CANADA

MEMBERS

Dr. Kenneth R. BOFF
Chief Scientist, Human Effectiveness
Air Force Research laboratory (AFRL/HE)
Wright Patterson Air Force Base 45433-7901
UNITED STATES

Mr. Paul CHATELIER
Potomac Institute/IDA
8021 West Point Drive
Springfield, VA 22153
UNITED STATES

Dr. Nahum D. GERSHON
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102-3481
UNITED STATES

Dr. Francois LESCREVE
Defence Staff - DGHR
Sectie Werving - Divisie Beleidsvoorbereiding
Kwartier Koningin Astrid
Bruynstraat, B-1140 Brussel
BELGIUM

Dr. H.J. WOODROOF
Programme Co-ordination Office
Defence Science & Technol,ogy Laboratory
Room 2, Building A1 - Dstl Fort Halstead
Sevenoaks, Kent TN14 7BP
UNITED KINGDOM

Prof. Dr. Paul NOJA
Gajon Institute of Technology
Via Paolo della Cella, 13
16135 Genoa
ITALY

Dr. Robert SEIDEL
1709 Belle haven Road
Alexandria, VA 22307
UNITED STATES

Dr. Bruno SICARD
Etat-Major de la Marine
Division “Programmes”
Antenne de Toulon - BP 55
83800 Toulon
FRANCE

PANEL EXECUTIVE

Col. Carel BANSE, MA
BP 25
92201 Neuilly-sur-Seine, FRANCE
Tel: +33 1 55 61 22 60/62   Fax: +33 1 55 61 22 98
Email: bansec@rta.nato.int or pelatd@rta.nato.int

Technical Evaluation Report

Robert J. Seidel, Ph.D.
1709 Belle Haven Road
Alexandria, VA 22307
USA
bsseidel@worldnet.att.net

1.0 INTRODUCTION

From 1985 through 1995, NATO held a series of workshops on technologies associated with military training. They covered Computer-Based Instruction, Training System Design, Distributed Training, and Virtual Reality. While technologies have dramatically changed, many of the problems and issues related to training technologies, its application, acceptance, and effectiveness are still with us today. The re-structuring of the military forces in size and composition, along with decreased funds for real exercises, has continued to focus attention on how to maintain a well-trained and flexible force. Recent changes in NATO and the European infrastructure have stimulated the RTO to sponsor a symposium focused on technologies applicable to improving military training.

Theme and topics

This unclassified Symposium was held in Genoa, Italy, from 13 to 15 October 2003, and was supported by the Human Factors and Medicine Panel of RTO. The Symposium audience included experts from NATO countries, PfP nations, as well as invited nations. The theme was to share data and establish working relationships that can continue to deal with the problem of training and simulation from a human systems integration point of view. The areas human factors of training systems as well as research in cognitive sciences all would be addressed**.** The symposium addressed the following topics:

1) **Technologies:** Intelligent Agents, Networked and Web-based Training, Virtual Reality, Embedded Training, Intelligent Training Systems, Multi-dimensional Interface Development;

2) **Training Applications**: Training System Development, Portable/Wearable Information Systems, Team and Group Training, Individual Training, Instructor Training, JOINT Service and Coalition Training, Situation Awareness, Decision making skills (How to accelerate tactical expertise development) (Critical thinking skills and the new partnership between man and machine), Embedded Training, Counter Terrorism, Distributed Mission Operations and Training, Training Value (Effectiveness), Value added, Cost effectiveness, Productivity;

3) **Cultural Issues**: Technology Acceptance, Coalition Training, Generational Differences, Instructional Design, Cross-Cultural Training of Instructors;

4) **Training Issues/ Problems:** Fidelity, Standards and Guidelines, Interoperability and Reuse of technologies, Multi-disciplinary Training Development Teams, Emerging roles of and demands on training and advanced training technology, Lessons learned about "what works" from existing technologies, and their promise for training today's military, The military value (cost-effectiveness) of advanced training technologies.

We know that military training means definite but very different things to different people. To the commander of a military unit, it means exercising troops in the field or sailors at sea so that they operate as an integrated, coordinated unit. To military personnel managers, it means preparing and certifying individuals across a full spectrum of occupational specialties that includes cooks, dog handlers, tank turret repairers, radar technicians, and fighter pilots. To developers and providers of major military systems, it means exercises performed in simulators or on the systems themselves. To all concerned it means preparing individuals drawn from a civilian society to perform as professional military personnel. It is distinguished from other forms of training by its emphases on discipline, just-in-time preparation, and the training of collectives.

Military training must prepare individuals to enter into harm's way and perform physically and mentally demanding tasks at the highest possible levels of proficiency in a time critical manner. This requirement may be the defining characteristic of military training. It can mean the difference between life and death.

Training places an emphasis on efficiency and achieving levels of knowledge and skill as quickly and inexpensively as possible. Training involves design, development, implementation, and evaluation. And these capabilities draw from research in many areas such as human learning, cognition, telecommunications, modelling, simulation, human factors and ergonomics, and advances in computer applications (Tobias and Fletcher, 2000).

Today's infusion of technology into nearly every aspect of military operations has significantly altered the nature of military engagements. It has increased the complexity of military operations, the number of tasks that individuals must perform, and the demand for knowledge and skill among military personnel. This complexity is compounded by the speed and mobility of modern military operations, the lethality and long reach of modern weapons, and requirements in modern doctrine for both dispersion and rapid composition of forces. In essence we can now see a closer relationship between human factors and training than ever before.

Symposium Highlights

There were four highlights to the symposium. The first consisted of the opening remarks by Dr. Dee Andrews, co-chairman of the symposium, and Admiral Accardo. They set the tone, the theme, and the challenge for the NATO R&D community. In the over 40 years during which time I have worked with technology in education and training, the overwhelming obstacles, or needs, which had to be served were not technological. Rather they were much more diffuse. They were individual, organizational, and institutional. The challenge from Dr. Andrews' talk is to share information amongst alliance and partner nations in NATO. The question posed by his talk, implicitly if not explicitly, is for the cultures of this multinational organization to go beyond its cultural, economic, business, and individual representations of borders, and develop a truly supra national organization and community that truly depends on each other outside of the military context. Such an organization would have to be open and could then be called, in the context of Andrews' speech, a learning organization. With the availability of advanced technologies and the opportunity for creating truly multinational synthetic training environments, the question for NATO to answer is this possible: can we do this, can we accomplish cross-cultural sharing of information, facilities, and training goals.

Admiral Accardo's video, depicted that at least among seven nations, the sharing of mission training rehearsal information and the creation of a combined training environment, with advanced technologies, and using an embedded training approach, could accomplished an effective joint training exercise. Indeed, his video demonstrated the implementation and operational setting of this viability.

Two presentations dealing with the Project First Wave illustrated the exciting potential to actualize the promise intrinsic to Dr. Andrews theoretical proposition. These presentations accomplished two things. First, they clearly identified the technical and technological obstacles that must be overcome, when asking several nations to share mission training exercises with simulators and various types of synthetic opposing forces (the interoperability issues). They also showed that these technological obstacles even in a preliminary attempt at orchestrating this mission training can be overcome by goodwill and cooperation among nations. On the other hand, they also highlighted the need for nations to overcome their insecurities and cultural differences concerning the sharing of information. In fact, the principal difficulty seemed to be, and I think will turn out to be, establishing a way that each nation can both share information and facilities with its fellow nations in NATO, and at the same time retain an acceptable degree of security for itself.

The third highlight was the research presentation concerning the evaluation of the virtual environment (VE) for aircraft maintenance training. The value of this study was in its careful attention to experimental detail in conducting research to evaluate both the effectiveness and the cost benefit of using the virtual environment for training. They carefully asked the question whether or not VE would be useful to solve a real training problem. They set about conducting a front-end analysis, completed a systems approach to training, and evaluated their product in the real world. In terms of its attention to detail, careful consideration of experimental design, and the concern for practical significance as well as statistical significance, this research project could serve as a model for other nations to follow.

Lastly, although only one presentation concerned itself with experimental validation involving the teaching of critical thinking to leaders, a number of papers concerned the need to shift from procedural kinds of training to a much more holistic, cognitive development approach for training individual leaders and teams in decision making. Ironically, the field of military training, which usually leads the field of education in applying innovations in items like technology, could benefit from the years of educational research on cognition and meta-cognitive development. In fact, one presentation specifically concerned the application of the theory of tacit knowledge, developed largely in educational environments, to military tactical decision making. The reason for the shift in emphasis to cognitive development seems to be the recognition that with the increasing complexity of using advanced technologies in cockpits, on shipboard, and in warfare in general, much more abstract, cognitive demands are placed on individual soldiers, leaders, and teams alike. The nature of likely conflict, high-intensity, frequent, and unpredictable location combines to demand, flexible, self-sufficient, and highly mobile forces. Together, advanced technologies and the changing nature of warfare place increasing cognitive demands on all personnel.

Symposium Program

Following two keynote presentations, the symposium was divided into seven sessions of various advanced technologies. Briefly, the sessions covered the following.

(1) **Simulation And Virtual Reality-** Examples of research were provided for land, sea, and air.

(2) **Research Results-** All of the efforts, with varied focus and interests showed an interesting potentially valuable area to improve NATO training. One focused on soldier values, another on the change towards holistic cognitive training for airmanship, and the third with limited support for a novel approach to measuring transfer.

(3) **Team And Collective Training-** A summary of this area in the symposium can best be characterized as requesting that training of leaders be changed to emphasize higher order thinking skills, with a focus on scenarios within a domain.

(4) **Training Analysis and Evaluation-** Two tools for aiding collection of reliable data during training were described in the context of feasibility testing. They still await complete evaluation. A third tool, a survey questionnaire and focus group technique, to aid technology insertion into

curricula was described. However, since it was developed in the context of a command-driven atmosphere, it awaits more study to see if it would work in other contexts.

(5) **Distributed Simulation-** Two presentations described a NATO first; i.e., multi-national mission training (7 nations) in a distributed synthetic environment. They identify the exciting potential and the problems of such a massive undertaking. A third presentation attempted to illuminate the problems of communicating with or without face to face possibilities with two nations involved. Both of these efforts can best be described as feasibility studies and are excellent examples of first steps in exploiting the potential of a multicultural, multinational training environment.

(6) **Command and Control/Leadership Training-** The use of technology-aided training in cognitive development for tactical leaders was the thrust of two of the three papers. Both approaches are still being evaluated; but in a preliminary examination, seem to be effective. The third presentation is an attempt to design a functional architecture for integrating synthetic environments into a crisis management capability for civilian emergencies.

(7) **Training Methods and Perspectives-** Perhaps the most significant presentation in terms of advanced technologies applied to training was given in this final session. This session actually presented what could be considered a model for how to do complete research from initial conceptualization through implementation and cost-effective justification. Another presentation pointed out the gaps in attempting a model the learner in a complex intelligent tutoring environment. A new application for an already proven technique, the systems approach to training, was provided in another presentation. Finally, an overview was provided of US Naval Research studies of advanced training technologies.

2.0 TECHNICAL EVALUATION

First, let me say what an honor and privilege it is to serve as the “technical evaluator”. It's amazing how the ambience of the majesty here in this hall and the magnificence of the products in the Tuscany countryside can add clarity to my own cognitive development. Having recognized that, I want to apply an historical perspective in organizing my remarks. I believe it was John Dewey around the turn of the 20th century, around 1910 or so, who put forth the notion that learning takes place based upon the actions and the transactions between the organism and his environment. In 1932, it was Bartlett who talked to us about the value of learning schemas in cognitive development. And in 1944, a perception psychologist named Egon Brunswik, provided the seeds for the notion of the importance of context and naturalistic observation. In fact, he found a reversal in the horizontal-vertical visual illusion in going from the laboratory to the real world. As a result of his research he coined the term "ecological validity". Now, somewhere along the way psychology got lost for awhile. Behavioral psychology, especially of the type that BF Skinner advocated, essentially overstated its case, and ignored what was going on inside the organism. Perhaps it is epitomized in the words of JR Kantor, Fred Skinner's mentor. I asked Professor Kantor at a colloquium, what he thought about Fred Skinner's book, The Behavior of Organisms. He mused for a moment, and then replied that Fred seemed to be a little bit immodest; and that he would have preferred that he use the title "Some Behavior of Some Organisms". I think that psychology has had this kind of problem for quite awhile of overstating and overreaching interpretations from a very limited paradigm. Examples of the value of functional context have existed for quite awhile in the military training literature; and I think is it was illustrated with the presentations in this hall as well. Another important constant in the psychological literature, regardless of theoretical allegiance, is the fact that feedback and reinforcement are significant factors in learning; and from a practical, military perspective, simulation as discussed in this symposium provides for an opportunity for almost immediate feedback and it is or can be related to the antecedent behavior to increase the speed of learning.)

I want to thank you all for responding to the questions, which I sent to all of you and to know that the purpose all of my correspondence was for formative evaluation; that is, to help you communicate more clearly to the audience. Today we are engaged in summative evaluation.

However, evaluation suggests that somehow I am going to provide a grade to each of you. Indeed that would be impossible since each of you are already among the top in your fields of expertise. Perhaps what I can provide is an assessment of how I see the overall program, an indication of how far we have come over the past decade, what it could mean to the training community, and provide a few comments where I believe your efforts could take you in the future.

So you see, I don't view this so much as evaluation as I do ~~as~~ a study in alignment. So, the focus is on alignment, alignment with the questions NATO might be interested in and how the value of the research reported here, might align with the needs of NATO in the general area of advanced technologies as they could be applied to training.

I was privileged to have served as chairman of a previous NATO research study group almost a decade ago on this same topic with many distinguished colleagues, some of whom are participating in this symposium. During this period, we summarized the lessons learned from our years of work and made a number of recommendations and offered a number of observations for guiding NATO training R&D for the future. It is with this background that I will enter into an elaboration of my impression of the general level of alignment of the current symposium with the outcomes of our series of workshops under the DRG and how I view the future of training technology and its impact on the alliance that we represent here today.

As an example, one of the most significant recommendations our DRG research study group made was to visit in depth three primary domains of training technologies: distributed learning, embedded training, and virtual reality. Each, we felt, warranted separate and detailed consideration because of their exciting potential for the NATO communities' training base in a time of shrinking budgets, smaller forces, and the need to provide flexible and agile training on demand in a cost-efficient manner. Clearly, the emphasis in the presentations at today's HFMP symposium stressed distributed learning and virtual reality, without much on embedded training. Perhaps, in the future there may be a merger developing between virtual techniques and embedded training. Such a possibility was indicated in Admiral Accardo's video. I would have liked to have heard more from the participants about this topic.

The following slides are included here rather than in an appendix to further clarify my points of the alignment of the past RSG with the current HFMP symposium.

As I indicated, this first slide addresses the specific issues which our RSG members believed needed to be addressed by future NATO study groups attempting to make recommendations to NATO for integrating advanced technologies into training:

This next slide now goes one level deeper into the specific technologies we covered and which align very closely to the content of this current HFMP symposium.

1. Training Individuals and Individualized Training.

Overwhelmingly, the literature supports tailoring instruction to the needs of each learner. This includes flexible time and place. The efficiency by using computer-based instruction, clearly shows at least a 30% savings in time. Therefore, intelligent tutoring system projects must be justified on a value added basis. ITS development demands a merging of AI with training strategies, task analysis, and multimedia; and therefore particularly emphasizes the need for multidisciplinary development teams. These efforts are neither inexpensive to obtain or maintain.

2. Architecture.

We focused here on that which can be affected by the developer of learning software or the user/manager of the software. Therefore, since we are nominally dealing with only software associated with training, the issues of re-use, shared use, and interoperability were the core concerns of our understanding of architecture. The architecture must be flexible and robust. Among other things, we advocated the development of a common taxonomy of characteristics for various shareable objects; so that they might be shared across training users with confidence, even though the objects were created by different builders. Recognize that this whole approach is iterative and can be hindered by premature establishment of fixed standards.

3. Authoring Systems.

Maintain an updated information base about currently available authoring systems. Maintain an updated set of evaluation criteria to permit the training developer to make systematic judgments about the types of authoring tools needed for their specific applications. Make sure that the final selection is based upon the capabilities required and the unique needs of a given situation.

4. Distributed Learning.

This flexible learning strategy provides learning on demand at what could be a significant cost savings since travel and the need for instructors is reduced. An interoperable network should be used for transmitting the variety of media and technology since it allows for constant injection of new and emerging capabilities. The costs of converting courses for use over a distributed learning system continue to be a significant funding challenge. And it requires that NATO provide consistent and meaningful cost and technological formulae to ensure interoperability as well as reliable, credible, and accountable data when planning and/or modifying distributed learning systems.

5. Virtual Environments (VE) and Training.

It was agreed that the current knowledgebase was limited then that the field of VR is quite dynamic. Questions arising concerning the application of various levels of VR capabilities in situation rehearsal and training suggest the need for exploration of many aspects of the development of the underlying technology evaluated in many different contexts of training. Similar to the concerns raised concerning intelligent tutoring systems, there is the question of when and where there is value added; i.e., when the number of different kinds of dimensions in VR and the VR itself provide a cost justification for its use. So there were many issues that we felt needed to be raised and cautions to be noted, when considering the application of VR to training.

I therefore offer the following cautions and guidance:

- DECIDE what the task is for the human,
- DETERMINE the level of reality required for training needs,
- IDENTIFY the appropriate interactions and interoperability required in the VE situation,
- EVALUATE the transfer to the real world environments after training in VE,
- DETERMINE the need for authoring tools,
- IDENTIFY the interdisciplinary requirements to do it right,

- GIVE the user sufficient time to experiment with the system (especially when it's complex),
- USE questionnaires after each application to gain information.

6. Cultural Needs.
Last but by no means least, we felt that the organizational as well as the varied national cultures, values, readiness to accept change, etc. all must be taken into account when considering implementation of new technologies. “Culture, we broadly define as the range of values and social norms of specific groups like instructor, trainees, officers, enlisted personnel, or of different NATO countries.”(p. 150, Seidel and Chatelier, 1997).

So, how did the presentations of this symposium address these issues? Three presentations (Foster and Fletcher, Chipman, and Wulfeck) explicitly addressed the notion of individualization of training. Kelly and Smith directly, and perhaps Foster and Fletcher implicitly were concerned with architectural issues. Four papers presented concerns for the development of tools to aid instructors or observers. However, nothing really was presented on authoring systems. Was there no interest? Are there no longer any problems in authoring? The overwhelming interest was on distributed training (nine papers.) and virtual reality and training (3-6, depending upon one's definition; i.e., total immersion versus virtual groupings through cyberspace.). With regard to cultural issues, there were five relevant, but only one which explicitly addressed it. Yet this topic continues to be of extreme importance as units and countries are tasked with accepting and implementing change with greater rapidity as technologies and the nature of warfare evolve; e.g., more self-sufficient, smaller units, with multinational composition, dealing with scattered high-intensity conflicts.

There is of course the ultimate question for policy makers, and a difficult, yet significant one for the researchers to address also. Does the product resulting from the current research provide sufficient value- added outcome to training in order to justify the cost of the investment in the product? It’s the old cost-effectiveness question that our former colleague, Dr. Jesse Orlansky, used to raise and make a lot of people uneasy. I call it, not cost-effectiveness, but rather the “tolerable cost” question. How much money am I willing to invest (cost to tolerate) for how much achievable incremental gain in training? And the question must be answered, not by end of training criteria, but rather by relating it to a transfer criterion of operational performance.

I would now like to address how well do the reports at this symposium deal with this and all of the questions posed? And how well does the symposium as a whole address the issues and questions raised by the earlier RSG on this topic?

1. Does the project being reported support the overall theme of the symposium? This issue was dealt with by the screening of the Program Committee. However, this could only be done on a limited basis since the Program Committee only looked at the abstracts and tried to project into the future about the papers.

To repeat the comments made above, three presentations (Foster and Fletcher, Chipman, and Wulfeck) explicitly addressed the notion of individualization of training. Kelly and Smith directly, and perhaps Foster and Fletcher implicitly were concerned with architectural issues. Four papers presented concerns for the development of tools to aid instructors or observers. However, nothing really was presented on authoring systems. Was there no interest? Are there no longer any problems in authoring? The overwhelming interest was on distributed training (nine papers.) and virtual reality and training (3-6, depending upon one's definition; i.e., total immersion versus virtual groupings through cyberspace.). With regard to cultural issues, there were five relevant, but only one which explicitly addressed it. Yet this topic continues to be of extreme importance as units and countries are tasked with accepting and implementing change with greater rapidity as technologies and the nature of warfare evolve; e.g., more self-sufficient, smaller units, with multinational composition, dealing with scattered high-intensity conflicts.

(The remaining issues I leave to the authors to consider in order sharpening the focus of the papers as they see fit for final submission to NATO/RTO.)

2. Is there a clear-cut military need to be served by this research or study? Is the need clearly articulated? Who were the users for the products of this research project?

3. What is the capability that will be provided by the successful completion of this project? What is the deliverable? Are the project research goals stated clearly?

4. Is the approach clearly articulated and consistent with research goals?

5. Are the results supportive of the military need?. And/or the research goals?

6. Finally, are the conclusions/recommendations justified by the study with respect to:

    a) the science,

    b) the military, and

    c) the policy?

Next, I will present highlights of the various session areas without referring to specific authors. Those commentaries, along with the three slides are included as appendices.

Simulation And Virtual Reality

Examples of research in these areas were provided for land, sea, and air. What was lacking in the presentations, and needs to be included in the next generation of studies are the use of a transfer criterion from VE to the real world, some focused and military relevant cost-effectiveness data, and data on the influence of multi-national as well as within nation cultural considerations all focused on transfer and learning effectiveness.

Research Results

All of the efforts showed an interesting and potentially valuable area to improve NATO training. The research studies were varied in their interests and in their focus on advanced training technologies. One study concerned an examination of the influence of individual values on soldier performance which is becoming very important as our role during combat changes. The second advocated a change in the approach of training, modern air crews towards a holistic, and constructivist, cognitive training approach. The third provided some data to support in a limited way the use of a backward transfer approach towards validating simulation based training. All require more comprehensive study in order to prove their points.

Team and Collective Training

A summary of this area can best be characterized as requesting that training of leaders be changed to emphasize higher order thinking skills, with a focus on scenarios within a domain. The research of Helsdingen, et al provides some support for this in the area of training for military tactical command. Cook presents a rational argument for it and Weeks provides a limited feasibility study in the area.

Training Analysis and Evaluation

Two tools for aiding collection of reliable data during training were described in the context of feasibility testing. However, from the presentation, it appears that they still await complete evaluation. A third tool to aid technology insertion into curricula in order to make the courses more effective was also described. Since the latter involved command selection of which courses should receive new technology as opposed to relative difficulty in terms of objective pass-fail criteria, the survey technique and results would seem to be of limited value. The approach needs to be tested in a curriculum-need driven context where the course data suggest that technology could have a positive effect on student learning by providing for example, more time on task for individuals with built in adjustment to each one's needs, or by adding simulation capability to improve transfer possibilities to the operational world.

Distributed Simulation

Two presentations described a NATO first, multi-national mission training (7 nations) in a distributed synthetic environment. They both identify the exciting potential and the problems of such a massive undertaking, the first on a global scale, and the latter, on some of the detailed problems already surfacing. Another presentation discussed issues involved in a more limited distributed simulation training environment with 2 nations involved. They attempted to illuminate the problems of communicating with or without face to face possibilities.

Both of these efforts can best be described as feasibility studies and are excellent examples of first steps in exploiting the potential of a multicultural, multinational training environment. Clearly, this cross-cultural perspective is realistic and extremely important for NATO, and requires support for further R&D.

Command and Control/Leadership Training

The use of technology-aided training in cognitive development for tactical leaders was the thrust of two of the three papers. Both approaches are still being evaluated; but in a preliminary examination, seem to be effective. The third presentation is an attempt to design a functional architecture for integrating synthetic environments into a crisis management capability for civilian emergencies. The purpose is to organize military command and control to provide assistance. The unique feature to emphasize here is they included all stakeholders in their initial analysis in an attempt to make their architecture as rich as possible. This looks promising as both a potentially useful architecture plus an example of how to incorporate different cultures into a development effort.

Training Methods and Perspectives

Perhaps the most significant presentation in terms of advanced technologies applied to training was given in this final session. This presentation actually presented what could be considered a model for how do complete research from initial conceptualization through implementation and cost-effective justification. They started by assessing through front-end analysis whether or not virtual reality would provide a value-added approach. From a complete systematic approach to training through the research they were able to show at the end cost-effectiveness data using immersive virtual reality. They justified the cost of VR in that it was a low-cost replacement for using actual equipment and still attaining the desired achievement. They are now considering for the future to improve upon their system by adding other senses and by studying how to reduce cyber sickness. The question is for the future does an increase in realism with improved images, haptic sense, and gloves, and more, then evaluated in a practical context, cost-justify the improvement.

Another presentation pointed out the gaps that exist in attempting to model the learner in a complex intelligent tutoring environment. It raised the question why haven't we gotten there yet? I answer perhaps it's that it is hard to develop; and relatedly, now may be the time since with greater emphasis on cognitive demands in military systems along with rapid response requirements, flexibility, self-sufficiency of soldiers, and the unpredictability of high-intensity small and frequent conflicts, a value-added justification may now properly emerge. Perhaps people may be more willing to put money into this type of the project. It was also noted that cultural issues might well play an important role in this type of investigation, although none of the models seem to have included consideration of other than military values except perhaps on a limited scale.

A new application for a proven technique, the systems approach to training, was provided in one presentation; that is, to apply it in the early stages of systems acquisition on a global level. But this was an application of an old technology not the development or application of an advanced technology.

Finally, an overview was provided of naval research studies on advanced technologies. In reporting on the use of embedded training in shipboard training the presenter emphasized the value of the military setting for developing state-of-the-art teaching and testing and noted that open research issues exist in developing natural language processing as well as cognitive modelling.

Before closing, I would like to refer to a lessons learned conclusion and recommendation made in the administrative domain by the RSG, which I chaired. We found that to be effective, collaboration requires frequent contact, a narrow focus on specific subject areas, and that this must occur over a number of years. Yet the procedures set up by the then Defence Research Group or DRG required a short-term focus, covering interesting topics and reporting on them. We recommended an alternative policy, which would permit on a case-by-case basis continuation of such study in depth, contingent on the potential of a given topic for increasing contribution to the NATO military. I raise this issue now because many similar interesting and worthwhile topics, stemming from the stimulating talks and discussions here, remain to be pursued in depth. I simply raise this is a question for the RTO; and ask, how will they proceed?

I will end this assessment with how I started. I do not wish to be presumptuous. However, the tone was set at the outset by Dr. Andrews for NATO/RTO to decide if it is a learning organization, and in this time of rapid change in need of cultural variables to be considered. Will there be more in-depth cultural studies? It is somewhat ironic in that NATO itself is a natural test-bed for this. Will it happen?

3.0 CONCLUSIONS AND RECOMMENDATIONS

It was gratifying to see that the focus in this symposium was not strictly on the technologies per se, but upon the value of the technologies for training. By far, the major emphasis in the research reported was on the value of advanced technologies to enhance distributed training. There was also a significant shift towards emphasizing the value of cognitive development, in all facets of training. Finally, there was a beginning of recognition that there is a need to consider cultural factors when considering multinational projects. While resulting in very interesting and stimulating discussion, most of the studies that were presented could be characterized as feasibility studies, or demonstrations, and not research. Secondly, the use of a transfer criterion, namely operational performance, in order to measure the value and cost-benefit of the new technologies such as virtual reality was present only in a couple of studies.

The first requirement for future symposia of this type is to make clear for the participants that they understand the requirement for research data to be presented. Some presenters were claiming in their defence that the instructions were not clear about this requirement. Coupled with this, it should be stressed that the criterion for evaluating the utility of the advanced technology in training must be the transfer value towards improvement of operational performance. In that regard, the research also needs to be evaluated in terms of its relative cost-benefit over such things as job aids or alternative methods of training. To help in this regard, it should be emphasized that incorporation of advanced technologies in training should be need driven. The complexities of current and future warfare indicate that in-depth study of various approaches to cognitive development training coupled with advanced technologies would seem to be a promising area for further research. Continuing research on virtual reality in various applications and with considering the numbers and types of sensory or cognitive involvement is also recommended, with the caveat that a cost-effectiveness criterion is still a bottom-line requirement prior to implementation. Lastly, a direct research attack on the contribution of culture as an important factor in NATO training should be strongly encouraged.

4.0 Appendix A. Detailed Session Commentary

Simulation and Virtual Reality Category

Examples of research were provided for land, sea, and air. What was lacking in the presentations, and needs to be included in the next generation of study are the use of a transfer criterion from VE to the real world, C/E data, and multi-national as well as within nation cultural considerations.

1. Cohn's presentation described a very thorough systematic approach using the SAT process. The driving force for his approach to using VR was the complexity of operational tempo and the need for greater cognitive decision-making, with less instructors, and the **opportunity to deliver training anytime anywhere with a virtual environment training system and thereby lower costs** with delivery of instruction. **However, detailed data are required to prove this point.**

2. The presentation by Goldberg on the use of VE in a dismounted combat environment sounded like a good approach; however the evaluation of effectiveness consisted of **perceptions of improvement and utility of the system** we still wait the study of the relationship between these perceptions and the actual performance. This is another example of the **need to insist upon a transfer criterion from training in virtual reality to the real world in order to evaluate its effect on for performance and to evaluate the relative costs against the relative gains provided.**

3. While another effort, by Rollesbroich, et al. described a project geared towards **evaluating collective training in a multinational environment, and implied recognition of the need to make certain that different nations would be able to work with one another,** a couple of items still remain ambiguous. There needs to be an emphasis on the study of cultural issues across nations in various training environments. Secondly, it is not clear whether the skills being trained for are skills for individual with performance in a collective environment, or a true dependency across individuals to learn, as well as, to perform collective skills. This needs clarification.

4. Another presentation, by Wulfeck, described a proven technological system which has improved training, ashore and collective training at sea. Measures of effectiveness with alternative approaches were not discussed during the presentation, apparently due to the classified requirements. **The question remains as to whether or not there is a cost benefit gain over alternative approaches. Although off-line it was stated that the data exist, this needs to be discussed more thoroughly in the NATO forum if security permits.**

5. The Loftin study concerned the use of a virtual environment for checkpoint training in peacekeeping operations was interesting in that it provided some meaningful research data apparently showing the value of VR training. **Unfortunately, we still need information concerning relative costs.** Is it conceivable that this achievement could be made with the desktop trainer alone? **Also, what is still required is a transfer criterion indicator** so that we have some idea as to whether or not training has a positive effect in the real world and not just in the environment of the virtual world.

Research Results

All of the efforts showed an interesting potentially valuable area to improve NATO training. The research studies were varied in their interests and in their focus on advanced technologies. One concerned an examination directly of the influence of values on soldier performance. The second advocated a change in the approach of training, modern aircrews towards and holistic, active and constructive cognitive training approach. The third provided some support in a limited way for the use of a backward transfer approach towards validating simulation-based training. All require more comprehensive study in order to prove *their points.*

1. The Avazashvili presentation was a specific attempt to examine issues of **values** and how they might influence stress tolerance. It was disappointing that the only study in the symposium that directly concerned itself with **culture** lacked any real consideration of detailed analysis. There is no information concerning the size of the correlation between values and stress for Georgian soldiers. Information on various variables that might account for differences need to be examined. However, the topic is an extremely important one and should be followed up. It would also be valuable to address how the issues uncovered might be influenced by soldiers working in more advanced technological settings.

2. The Ebbage and Spencer presentation highlighted an important point that was raised by Dr. Andrews in his opening comments. They advocated a change in training for modern aircrew's airmanship. This would require that the UK see itself as a learning organization and therefore be open to the authors' advocated holistic approach to airmanship training. The technology that might be used to implement this new approach could be desktop trainers but as the authors emphasize the main point is **to change the approach towards active, constructive, cognitive skills training**. One might view this approach as technology in training being the servant of, and not the driver of change.

3. Goettl, et al described a simulation environment for visual threat recognition and avoidance which was validated by using a "backward" transfer approach in which experts performed on the training system in comparison with novices or less expert personnel. **The data showed a significant backward transfer effect.** Although correlations for the most part dealing with total mission, total hours, or threats at one's own ship, at most accounting for 4 percent of the variance, perception of threat at other aircraft was the only significant correlation and accounted for between 16 to 20 percent of the variance. **While encouraging, these data provide limited support for the approach, and it needs to be re-examined and validated in a larger research effort.**

Team and Collective Training

A summary of this area in the symposium can best be characterized as requesting that training of leaders be changed to emphasize higher order thinking skills, with a focus on scenarios within a domain. The research of Helsdingen, et al provides some support for this in the area of training for military tactical command. Cook presents a rational argument for it and Weeks provides a limited feasibility study in the area.

1. van den Bosch, Helsdingen, and de Beer. This presentation emphasized that in our new complex world environment, training for military tactical command requires recognition that problem solving skills can be taught and that we can learn from experts. The expert typically has a bigger database for the domain, and novel situations are handled differently than for novices the expert's looks for the pattern gaps and consistency and they are able to provide metacognition and be critical of their own behavior. The novices don't see patterns and they stick with the first strategy from isolated piece of evidence. Therefore their suggestion is to adapt the training of leaders to naturalistic kinds of strategies and decision-making. Training in the use of these scenario based exercises leads to better process development in terms of argumentation for situation assessments etc. as well as outcomes. It would be interesting to characterize this in my term, near-term transfer. I would submit however the epitome of training for role of decision-making and complex environments can be provided through training involving multiple examples multiple contexts and ultimately lead to far-term transfer or the ability to see relationships in other domains as well as the one focused on in a particular training environment. **I guess the question that I have is what is new? And the answer I come up with I suppose is the domain of application being tactical decision-making as opposed to for example troubleshooting as well as possibly the extent of the implementation. By implication the argument is made for training metacognition and**

other self-monitoring skills.

3. The Week's study is essentially a feasibility study. It's not clear what is unique. It provides for command and control training at different levels. What he had was an acceptance test for the value of the computer program which he used as validity. The instructors were able to learn to use the course for trainees. We need a transfer criterion. The overriding need to be investigated is the trade-off between job-aiding and training.

4. Cook's presentation was a broad theoretical treatise, providing yet another argument for changing training into a critical thinking, cognitive development, decision-making and problem solving environment. He advocated, because of the new cognitive demands, using meta-cognitive and insight development training. The driving force for this new environment of training is the new environment for warfare: with rapidly changing enemies you need to have rapidly changing strategies therefore you cannot have procedural training by itself. Without mentioning that, he also supports the notion of a spiral curriculum such as I've advocated my book. **He, like Helsdingen, also goes only as far as near-term transfer in his advocacy for cognitive training. While not particularly new, that is, the cognitive field took off in psychology during the 1960s, it is an example perhaps that the real world, the applied world of the military, is finally ready to embrace more complex requirements being matched to more complex cognitive training. It is an example perhaps that the real world, the applied world of the military, is finally ready to embrace more complex requirements being matched by more complex cognitive training.**

Training Analysis and Evaluation

Two tools for aiding collection of reliable data during training were described in the context of feasibility testing. They still await complete evaluation. A third tool to aid technology insertion into curricula in order to make the courses more effective was also described. Since the latter involved command selection of which courses should receive new technology as opposed to relative difficulty in terms of objective pass-fail criteria, the survey technique and results would seem to be of limited value.

1. Allen's presentation describes the development of a handheld personal data assistant. Unfortunately the study was carried only so far as to show that it was usable and that redesign recommendations were suggested. **What is important and what Dr. Allen discussed in off-line communication is that an answer to my e-mail question concerning the need for a target criterion, he agreed that we do need to have some idea of a target criterion for this device in terms of the errors that we are willing to consider acceptable and how this compares with previous devices or methods in terms of their error rates. So it could be a useful tool for performance measurement aboard ships but we await the final evaluation.**

2. van Berlo, et al. also developed a data measuring system for standardizing consistent feedback across teams of observers, during distributed training, implemented on a handheld device. **Again what was reported was a feasibility study** of this tool and as the author noted "further research will be conducted on the prototyping on the validation of the tool."

3. Capt. Sidor, et al. described in interesting technique for determining the need for training technologies to be inserted into a curriculum. While they assert that the project was need driven, they worked with pilots as subject matter experts in their survey. They also brought instructors into the laboratory. My question arose **why didn't they get performance data from the various curriculum outputs. This is what we did when we worked with the army schools and therefore found the needs driven approach much more viable by looking at output data in an objective manner; i.e., pass-fail rates in various courses. In off-line conversation, Capt. Sidor noted that the selection was really command-driven concerning which courses would receive the technology. This is an efficient way, but not necessarily**

effective , way to go about doing this.

Distributed Simulation Overall

Two presentations describe a NATO first, multi-national mission training (7 nations) in a distributed synthetic environment. They both identify the exciting potential and the problems of such a massive undertaking, the first on a global scale, and the latter, on some of the detailed problems already surfacing. Smith and McIntyre discuss issues involved in a more limited distributed simulation training environment with 2 nations involved. They attempt to illuminate the problems of communicating with or without face to face possibilities. Both of these efforts could be described as feasibility studies and are excellent examples of first steps in exploiting the potential of a multicultural, multinational training environment. Clearly, this cross-cultural perspective is realistic and extremely important for NATO, and requires support for further R&D.

1. Paper by Tomlinson, and van Geest on Exercise First Wave illustrates both the exciting potential for multinational training exercises, as well as the tremendous challenge in coordinating and sharing of information and facilities. While still in the planning and preparation phases, the authors report on significant lessons learned in these areas already by the seven participating nations. This truly represents a first for NATO, providing an opportunity for collective mission training. The target to conduct the training is September 2004. The presentation is comprehensive in that the authors show awareness of all of the potential problems that may be incurred, from equipment interoperability to security of information. We look forward to a successful conclusion of First Wave. Next year.

2. Cerutti and Greschke provide a second description of this **multi-national cooperation** effort on performing mission rehearsal training in a complex synthetic environment. **Theirs is a more "a nuts and bolts description". Yet it also emphasizes that the exercise is an excellent first step; but it is only that—it is a technical demonstration. In that sense, it is another feasibility study.** While Cerutti notes that their study, First Wave, was confined to mission exercises, I would assert that **we need to study in much more depth cross-cultural and organizational issues in order to implement and then feed back to the R&D community new requirements.** As one indicator of new requirements, the task was underfunded in that each country provided its own support, thereby creating interoperability problems since they had to patch together an ad hoc way to share data. Thus, some of the critical potential difficulties noted by Tomlinson and van Geest have been occurring already.

3. Smith and McIntyre in their study illustrated the differences and problems one faces in a distributed environment where no eye contact or nonverbal communication is available. They describe the first experience within this distributed mission training environment as negative. It must be noted however that until a group gets used to a system there will be skepticism and difficulties in using it. Secondly, only one of the groups was in a distributed environment, the U.K. and the other was co-located (the Mesa AZ group). There were also technical problems and these can easily create annoyances and feelings of dissatisfaction with the process. **The next study should involve both, multinational groups all of whom are in a distributed environment compared with multinational groups all of whom are co-located in order to properly make comparisons.** There were also technical problems and these can easily create annoyances and feelings of dissatisfaction with the process. **The next study should involve both, multinational groups all of whom are in a distributed environment compared with multinational groups all of whom are co- located in order to properly make comparisons.**

Command and Control Leadership Training

The use of technology-aided training in cognitive development for tactical leaders was the thrust of two of the three papers. Both approaches are still being evaluated; but in a preliminary examination, seem to be effective. The third presentation, by Kelly and Smith, is an attempt to design a functional architecture for integrating synthetic environments into a crisis management capability for civilian emergencies. The *purpose is to organize military command and control to provide assistance. The unique feature to emphasize here is they included all stakeholders in their initial analysis in an attempt to make their architecture as rich as possible. This looks promising as both a potentially useful architecture plus an example of how to incorporate different cultures into a development effort***.**

1. Lussier and his study showed that adaptive thinking training works for training tactical leaders. The use of a computer-based version of this training proved to be feasible and that students increased their performance during the time of training and were able to become more expert. The perceptions of the course were favorable by both instructors and students. Note that this in itself is a cultural indicator that the new computerized version would be acceptable. It is likely that the computerized version of the course will therefore be proved to be at least as successful if not more so than the traditional already implemented course. **The good news about the Lussier study is that they have an implementation link to the tank training school; the bad news is that a transfer criterion is still needed in order to next determine full-scale use and value. The premise behind this study is the need for training leaders to be flexible expert tankers, to be able to use knowledge rather than acquire knowledge. This is another paper that emphasizes the development of schemas or higher order thinking as an important advance over previous training.**

2. Kelly and Smith described a very important feature which seems to be missing in most of the studies reported on at this conference. That is, in order to develop an adequate functional architecture for training command and control regarding military assistance to civilian emergencies, they purposely **included all stakeholders** that might be concerned with how to provide adequate solution to crises. I was impressed with the detail as well in their data collection efforts. As noted by Mr. Kelly in correspondence the focus of their development effort is “on applying for integrating technology to provide training solutions” not on the pedagogy that might be involved. They also provided an important example of being able to use proven technology as part of their development effort (MILES) and therefore show how you might integrate new technology and old technology in an efficient way.

3. Psotka provided yet another example of **emphasizing cognitive development** as important in military leadership training. In this case what he was doing more as showing how you could create an automated self-development tool and base it upon science of the development of tacit knowledge. The experimentation reported showed that the online environment was superior to a paper and pencil environment. Not provided however, were the cost implications for implementing this innovation. These need to be provided. **Perhaps the most interesting part of the presentation by Dr. Psotka was that in his presentation he used metaphors from two different domains. His metaphors concerned Goldilocks and Tolstoy; yet his task for the audience was to transfer or transition the meaning of his metaphors into his talk’s target-domain of tactical decision-making. I would submit therefore that this presentation provides an excellent example of far-term transfer or cross- domain transfer, and I would submit, needs to be taught in leadership development. Moreover, if we are to train for flexibility, self-sufficiency, and adaptability, and a small or more mobile fighting force, cross-domain decision-making and critical thinking (far- term transfer) as well as within domain cognitive development (near-term transfer) needs to be taught to our Armed Forces.**

Training Methods and Perspectives

Perhaps the most significant presentation in terms of advanced technologies applied to training was given in this final session.

1. The presentation by Lieut. Purtee and Ms. Graci actually described what could be considered a model for how do complete research from initial conceptualization through implementation and cost-effective justification. They started by assessing through front-end analysis whether or not virtual reality would provide a value-added approach. From a complete systematic approach to training through the research

they were able to show at the end cost-effectiveness data using immersive virtual reality. They justified the cost of VR in that it was a low-cost replacement for using actual equipment and still attaining the desired achievement. They are now considering for the future to improve upon their system by adding other senses and by studying how to reduce cyber sickness. The question is for the future does an increase in realism with improved images, haptic sense, and gloves, and more, then evaluated in a practical context, cost-justify the improvement.

2. Foster pointed out the gaps that exist in attempting to model the learner and a complex intelligent tutoring environment. He raises the question why haven't we gotten there yet? I answer perhaps it's that it is hard to develop; and secondly, now may be the time since with greater emphasis on cognitive demands in military systems along with rapid response requirements, flexibility, self-sufficiency of soldiers, and the unpredictability of high-intensity small and frequent conflicts, a value-added justification may now properly emerge. Perhaps people may be more willing to put money into this type of the project. It was also noted that cultural issues might well play an important role in this type of investigation, although none of the models seem to have included consideration of other than military values except perhaps on a limited scale.

3. A new application for the systems approach to training was provided in the Verstegen presentation; that is, to apply it in the early stages of systems acquisition on a global level. **But this was an application of an old technology not an advanced technology.**

4. Susan Chipman provided an overview of the office of naval research studies on advanced technologies. In reporting on the use of embedded training in shipboard training she was able to emphasize the value of the military setting for developing state-of-the-art teaching and testing. She also pointed out that open research issues exist in developing natural language processing as well as cognitive modeling.

Advanced Training Technologies And Their Impact On Human Performance Improvement

Dee H. Andrews, Ph.D.
Senior Scientist
U.S. Air Force Research Laboratory
6030 S. Kent St., Mesa, Arizona, USA, 85212-6061

Summary

This NATO symposium is focused on advanced training technologies that can have a significant impact on the way NATO member nations prepare their personnel for peace and wartime duties. There have been some remarkable developments in our ability to better train personnel. However, merely improving training is not enough to assure that our military organizations are optimized for their duties. We must help our militaries to become "learning organizations" that wisely make use of a variety of technologies, including training, as we seek to optimize human performance improvement. Conceptual underpinnings that are requisite to forming learning organizations are examined. Strategic planning tools, knowledge management systems, and the Learning Management Maturity Model are among the topics explored. Those who seek to transform our militaries into learning organizations should view training and training technologies as two important arrows in the entire quiver of human performance improvement technology.

Introduction

Training is a critical part of NATO military preparedness. NATO militaries expend millions of hours in training each year, and some would say that virtually all non-wartime military activity could be called training. NATO military training spending easily exceeds a hundred billion U.S. dollars each year. In this advanced training technology symposium significant new advances are discussed in training technologies and methods. These advances will help to dramatically improve our capability to train individuals and teams to new levels of competence in a variety of warfare and peacekeeping domains. The papers in the symposium span a wide gamut of training topics including;

- infantry training
- peacekeeping training
- airmanship and aircrew training
- command and control training
- security training
- sonar training
- civil emergency training

Modeling and simulation, instructional and cognitive psychology, and educational technology are just a few of the tools that are being used to address the training domains. All of the papers provide a useful set of snapshots of the state of the art in training development.

This paper posits that no matter how effective we make our training technologies and techniques, we will ultimately fall short of our goal of optimal military performance if training is the only

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held in Genoa, Italy, 13 – 15 October 2003, and published in RTO-MP-HFM-101.

tool that we use. Training, as important as it is, should be viewed as only one arrow in our quiver of arrows that can be used to improve human performance in our militaries. I believe it can be shortsighted to assume that because we can perfect our training approach that we should assume that we are doing all we can to optimize the performance of our military personnel. I believe the time has come to move into a new era of human performance technology as we seek to make our militaries the "learning organizations" that they should become. I view training as a tactical tool, that when used appropriately with other tactical tools for performance improvement can help us to reach optimum performance in our militaries. To do that, we must view the performance improvement enterprise from an organizational strategic level.

The business community has been facing this issue for years. Try as they might, as spend as much as they could, our colleagues in the business community have gradually come to the conclusion that merely providing more and better training will not by itself produce the performance gains that businesses need to stay competitive. In many cases training had become an end in itself and was frequently not well tied to the overall strategic thrust and vectors that businesses were pursuing. Training departments in business were finding themselves on the outside looking in as businesses pursued new directions with which traditional training departments were ill suited to cope. The end result was that training departments were finding themselves marginalized as they had difficulty providing satisfactory answers to "Return on Investment" questions that senior business leaders were asking about the training enterprise.

O'Driscoll (2003) stated well the conundrum that training managers face:

"Today the training function finds itself at a crossroads. The unwillingness of training directors to let go of the bureaucratic machine they have created for themselves has become an issue that could call into question the value of training. In continuing to focus on measures that do not correlate well to organizational performance, training directors might in fact be compromising the future of their own functions." p. 8

Human Performance Technology

As our NATO militaries actively take on the quest of transforming themselves to meet the challenges of the post-cold war era, and the increase in international terrorism, it is vital that the role of human performance improvement be re-examined. Key questions for the training and training research community should be, "Will excellent training by itself be enough to produce the competencies required?" and "How can we help our militaries take a more strategic view of human performance improvement, of which training is an important tactical tool?"

I believe that Human Performance Technology (HPT), of which training is a piece, is the answer to the performance challenges our militaries face. What are these HPT technologies? Here is a partial list. They include both instructional and non-instructional interventions.

Examples of Performance Interventions of an Instructional Nature

- Classroom Instruction
- Small-Group Activities
- Video-Based Instruction
- Computer-Mediated Instruction
- Printed Self-Instruction
- Resource Management Systems

- Structured On-the-Job Training
- Distance Education Systems
- Accelerated Learning Systems

Examples of Human Performance Interventions of a Non-instructional Nature

- Organization Design (changes in the basics of an organization's processes – strategy, structure, systems, competence, and culture)
- Culture Change (response tendencies or behavior patterns that characterize people within an organization)
- Strategic Alignment (crucial organizational systems behind a common purpose or mission statement)
- Personnel Selection
- Motivational Systems (set of tactics and strategies designed to stimulate and sustain appropriate levels of goal-directed effort and affect.)
- Feedback Systems (information required for competent performance can be improved: through data manipulation, to effectively improve performance; and through improvement in the ways in which performance is directed or guided.)
- Incentive Systems (all rewards and remuneration given to an employee to elicit, improve, and maintain work performance.)
- Minimalist Documentation (reducing the obstacles to self-directed discovery and achievement that can inhere in modern systems and documentation.)
- Ergonomic Performance Aids
- Expert Systems
- Off-line resources (knowledge management)

The training community is of course quite familiar with all of the instructional interventions. The papers presented at the Symposium address all of the instructional interventions I have described, plus a variety of others. Decades of experience with these interventions have allowed dramatic increases in military performance because of better training.

Most of us in the training community are also familiar with at least most of the non-instructional interventions. These interventions come to us from a variety of disciplines (e.g., management science, industrial/organizational psychology, organizational development, human factors). For a more complete discussion on these non-instructional interventions please refer to the "Handbook of Human Performance Technology" (Eds. Stolovich and Keeps, 1999).

Rosenberg (2003), in Figure 1, graphically portrays the relationship of training and non-training interventions within the HPT tool set.

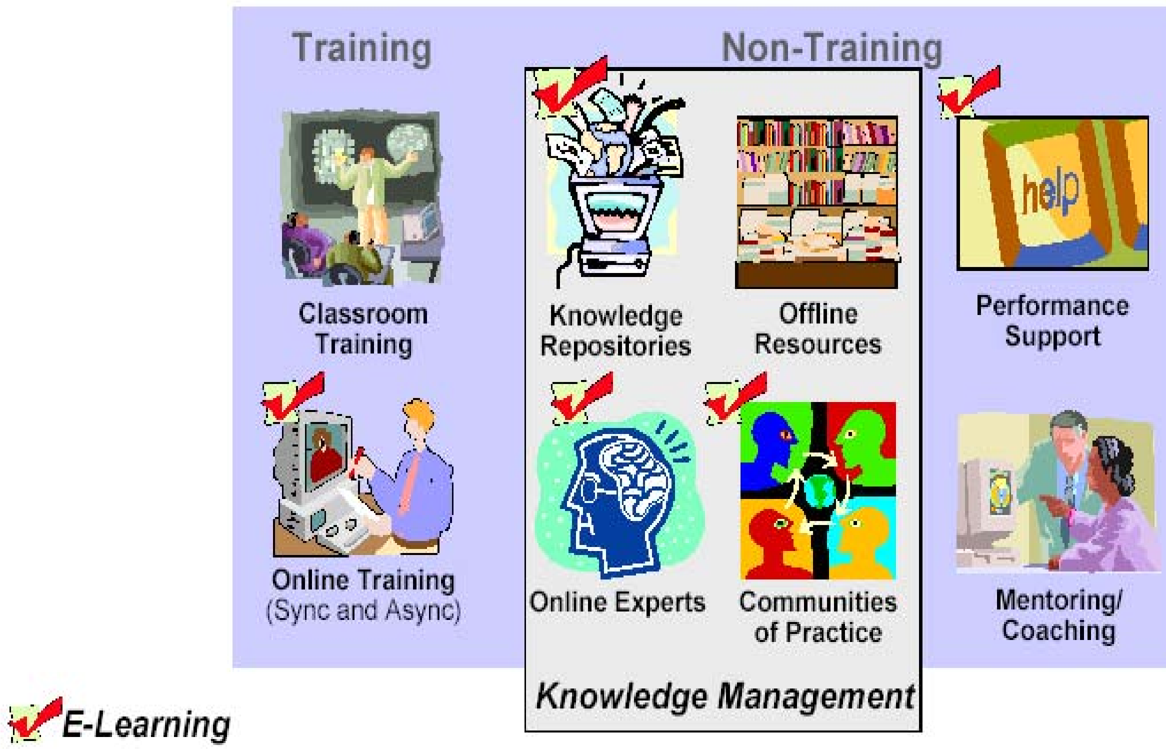


Figure 1: Relationship of training and non-training interventions (used with permission)

Learning Organizations

To create learning organizations in our militaries we must move from thinking of performance improvement as something that comes primarily from formal training, to viewing learning as a strategic approach that is a vital part of all a military does. Rosenberg (1996) describes a set of five transitions requisite to establishing a learning organization.

From	To
Learning as an end in itself	Valued performance as the primary measure of effectiveness
Training interventions as tactical responses to somewhat larger tactical problems	Performance technology as a strategic response to strategic needs relating to people and productivity

A view of training as overhead and support, susceptible to budget cutting and downsizing	Performance technology as a competitive resource, perhaps even more important during business downturns
Interventions placed in Human Resource functional chimneys that do not integrate with one another	An integrated performance improvement system that is systemic throughout the organization
A focus on educational results, for example, learning	A focus on organizational learning and business results

Brethower (2003) describes the goals of both Instructional Systems Development (training) and Human Performance Technology. As you will see below, he indicates that training has an essential role to play in improving human performance. However, when you compare the goals of ISD (training) to the goals of HPT it is apparent that HPT has a broader strategic reach. This broader reach is what is required to evolve an organization into a Learning Organization.

"Instructional Systems Development, used competently, assures that training delivers:

- the right knowledge
- to the right people
- at the right time
- to help improve workplace performance
- related to a significant performance issue

HPT, used competently, adds value by improving significant performance at the individual, process, and organizational level.

HPT, used wisely, adds value to an entire value chain: the organization and its customers, suppliers, employees, and financial supporters, as well as the physical, social and cultural environment." p. 12

Learning Technology Landscape

Addison's (2003) depiction of a "Learning Technology Landscape" (Figure 2) helps to put HPT into a broad context, starting with the society in which the organization exists and working down to the individual and team level. NATO militaries are part of the societies of their respective countries. The Learning Technology Landscape helps military trainers and researchers view the broader context of their efforts as they attempt to improve the performance of their militaries. It frames the key environmental levels, principals, and phases of HPT according to the systems approach to problem solving. As one examines the landscape cube, one can see that training may be useful in many of the cells of the cube. The overall organizational effectiveness of the training can be better understood by determining which of the cells training impacts. However, there will invariably be cells of the landscape in which training may not alone be effective for meeting the goals and purposes of the cell. In those cases, the HPT analyst should examine non-instructional interventions that can be used either in conjunction with a training intervention, or by themselves.

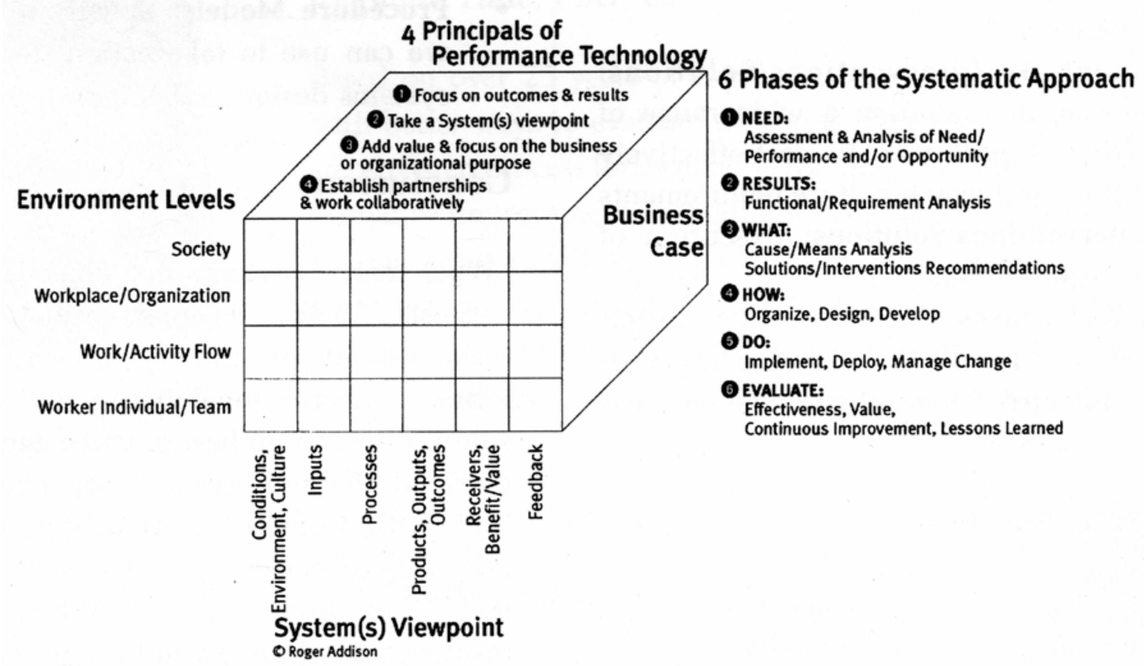


Figure 2: Learning Technology Landscape (Addison, 2003)

Addison explains that HPT methods and applications contribute to accomplishing one or more of the following:

"- Identification of Value: Clarification of the problem, organizational issue, or opportunity
- Definition of Outcomes: Specification of the requirements to create the value or close the gap between existing and desired results.
- Analysis of Performance: Identification of the factors in the performance system that can influence the production of outcomes to meet requirements.
- Selection of Interventions/Solutions: Choosing from a range of possible HPT applications those that will best meet the requirements, given the information about outcomes and performance
- Design/Development of Interventions/Solutions: Preparation for execution including a wide range of decisions about what it will take to cost-effectively implement the solution and match culture requirements.
- Deployment of Interventions/Solutions: Execution of the design to meet requirements
- Evaluation of Effectiveness: Measure intervention processes, outcomes, and results to determine how well they meet the requirements and what might be required to further improve results." (Addison, 2003, P. 14)

Knowledge Management

The last few years have seen dramatic growth in tools and databases that allow individuals and teams to dramatically access to vast stores of knowledge which would have previously taken years of research to find. The internet is the most significant, although not the only example. Rosenberg (2001) tells us that, "Knowledge management supports the creation, archiving, and sharing of valued information, expertise, and insight within and across communities of people and organizations with similar interests and needs." (p. 66). He tells us that far more than simply being a set of tools to manage information, knowledge management is about using knowledge to fulfill strategic organizational goals. He presents three levels of knowledge management.

Knowledge Management Level 1: Document management:
- access and retrieval,
- documents stored online

Knowledge Management Level 2: Information creation, sharing and management:
- capturing and distributing expert stories
- real-time information management
- communication and collaboration
- new content creation

Knowledge Management Level 3: Enterprise Intelligence
- leveraging organizational "know-how"
- performance support
- interacting with operational databases
- building expert networks

In addition to providing better training methods and technologies, we in the research community should be helping our militaries move toward Level 3 if we hope to have our militaries become true learning organizations.

This ability to systematically manage knowledge is becoming a key HPT tool. Rosenberg (2001) in Figure 3 presents a informative depiction of how the related constructs of "training", "knowledge management" and "performance support" can interact to form a strong HPT toolset. All three constructs are crucial in building a learning organization.

	Training	Knowledge Management	Performance Support
Purpose	Instruct	Inform	Guide Performance Directly
Stop Work?	Requires Interruption of Work	Less Work Interruption than Training	Integrated Into Work Tasks
View of Learning	Program Dictates How Learning will Take Place	User Determines How Learning will Take Place	Learning is a By-product of Performance
Goal	Transfer Skill and Knowledge	Resource for User; Catalyst for Organizational Learning	Assist Performance or Do It Completely

Figure 3: Relationships among training, knowledge management, and performance support HPT tools (used with permission)

Learning Management Maturity Model

Moore (2002) has developed a learning management maturity model that gives us a vision of what is required for an organization to truly become a learning organization. He indicates that there are five components of the learning management model.

- Focus on Strategic Impact of Learning
- Cultivate a Continuous Learning Culture
- Leverage Multiple Learning Channels
- Create Compelling Content
- Show the Value Back to the Organization

Figure 4 shows the stages of the model:

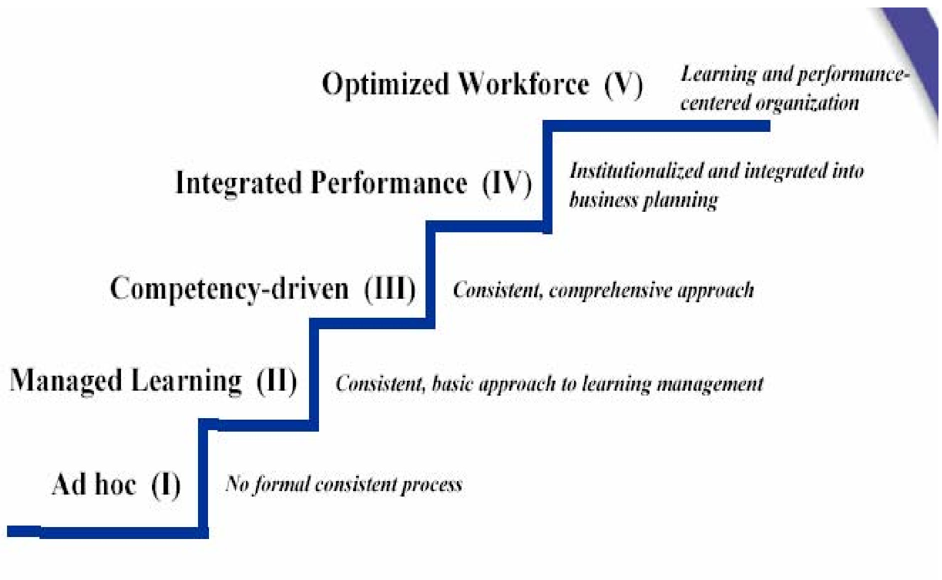


Figure 4: The Learning Management Maturity Model (used with permission)

Figure 5 depicts the nine key technologies and related pieces of the model.

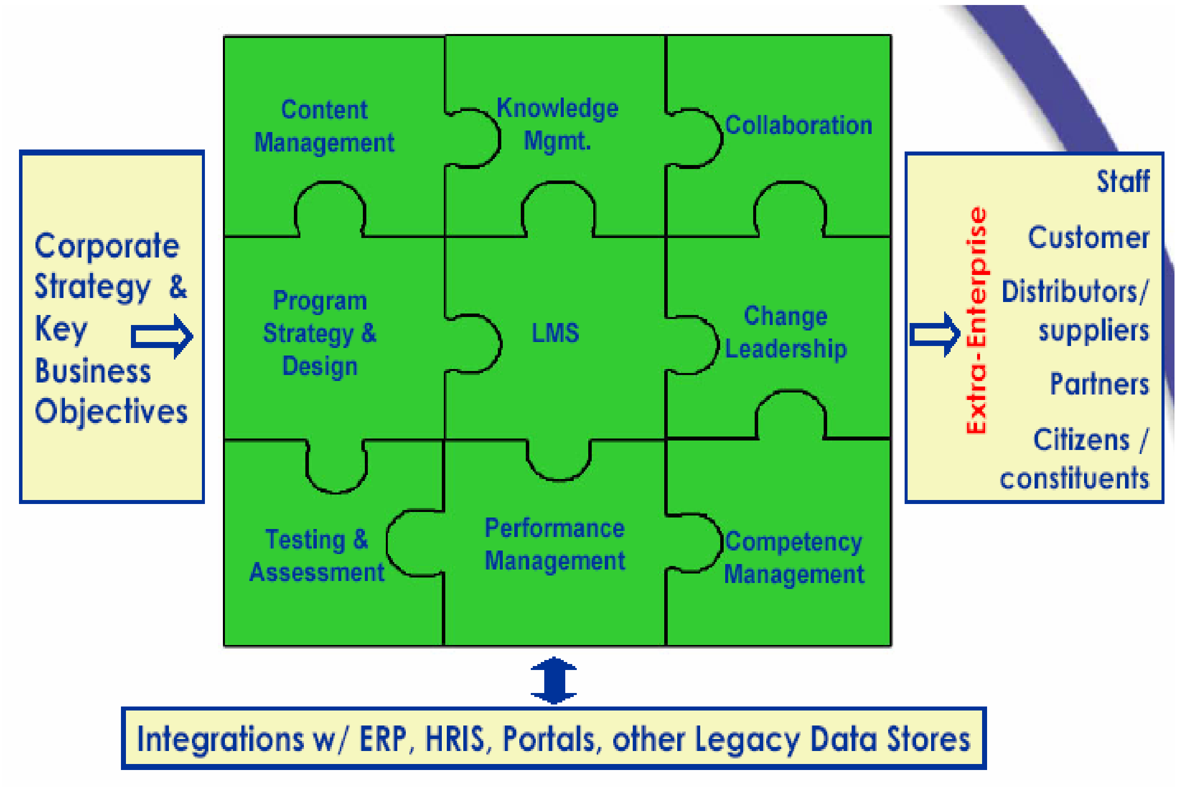


Figure 5: Nine Key Technologies and Related Pieces of the Learning Management Maturity Model (used with permission)

Content Management
- Authoring Tools
- Commercial Off-the-Shelf Providers
- Third party developers
- Learning Content Management Systems and Content Management Systems and Decision Management Systems

Testing and Assessment
- Kirkpatrick evaluations
- Pre- and Post- Testing
- 360 degree feedback

Knowledge Management
- Repository management
- Workflows

- Electronic performance support systems
- Advanced search and navigation

Learning Management Systems

- Catalog administration
- Event scheduling
- Compliance management
- Job development
- Resource management
- Skills and competency enablement

Performance Management

- Goal management
- Performance evaluations
- Talent management and Succession planning

Collaborators

- Virtual classrooms
- Blended learning
- Mentoring
- Discussions
- Online Meetings

Competency Management

- Skills libraries
- Competency maps
- Services
- Enablement tools
- Development planning

Moore indicates that the organization must have all nine pieces working synergistically together in order to reach the fifth stage of the model, the optimized workforce. He goes on to say that in his experience of consulting with numerous for-profit and government organizations he has never seen any organization get beyond stage three, competency driven. In fact seldom does he find organizations beyond even stage two, managed learning. That is, he sees the rudiments of a consistent, basic approach to learning management, but that approach is not comprehensive across the organization.

Moore is careful to point out that reaching stage four or five in the learning management model might well be a goal for many organizations, but should probably not be a goal for all organizations. It takes a considerable investment in commitment, resources and time to reach stage four or five. While the end result will be a more effective and efficient organization, the organization's leaders must be totally committed to the investments that must be made. Otherwise the organization and people will be frustrated by less than the full commitment.

How Training Technologies Can Help in Improving Human Performance

As mentioned above, training technologies are clearly a component of the instructional intervention part of the human performance improvement toolkit. Training is essential in imparting key knowledge, skills and abilities to an organization's personnel. The papers in this advanced training technologies symposium will describe a tremendous capability for instructing NATO's military personnel. However, in this paper I have described the need for a larger strategic view of Human Performance improvement, of which training is a key part.

References

Addison, R. M. "Performance Technology Landscape" Performance Improvement. Vol. 42, Number 2 pp. 13-15, 2003

Brethower, D. The Human Performance Technology Value Proposition. Performance Improvement, Vol 42, Number 2, Feb. 2003

Moore, C. Learning management maturity model: Five stages of learning maturity within the enterprise. 2002. Chris.Moore@THINQ.com

O'Driscoll, T. Improving knowledge worker performance. Performance Improvement, Vol. 42, Number 4. 2003

Rosenberg, M.J. (1996) Human performance technology. In R. Craig (Ed.), The ASTD training and development handbook (pp. 370-393). New York: McGraw Hill

Rosenberg, M.J. (2001) E-Learning: Strategies for delivering knowledge in the digital age. McGraw Hill: New York.

Rosenberg, M.J. (2003) Building management solutions. Presented at the Training Director's Forum, Phoenix, Arizona, June, 2003

Stolovich, H.D. and Keeps, E.J. (Eds.) Handbook of Human Performance Technology – 2nd Edition. Jossey-Bass: San Francisco, 1999.

Training Dismounted Combatants in Virtual Environments

Stephen L. Goldberg
U.S. Army Research Institute
12350 Research Parkway
Orlando, FL 32826
United States

Bruce W. Knerr
U.S. Army Research Institute
12350 Research Parkway
Orlando, FL 32826
United States

James Grosse
U.S. Army RDE Command
12350 Research Parkway
Orlando, FL 32826
United States

Summary

The U.S. Army Research Institute, U.S. Army Simulation, Training and Instrumentation Command, and the U.S. Army Research Laboratory recently completed a four-year effort to improve capabilities for dismounted soldier simulation. With increased emphasis on Military Operations in Urban Terrain (MOUT) new flexible training methods are needed that can represent a range of urban environments. Virtual environments have that capability but technological challenges must be overcome to produce an effective virtual training system for the dismounted combatant. This paper reviews improvements made during the project in the areas of representing the dismounted combatants' environment, producing realistically performing Dismounted Infantry Semi-Automated Forces, and developing of an After Action Review system that captures performance in MOUT. At the end of each year of the project a Culminating Event was held during which improvements made in technologies made during the year were integrated and then evaluated by soldiers. The paper also contains a discussion of soldiers ratings of the usability and training effectiveness of these capabilities. Significant improvements have occurred as a result of this project taking virtual dismounted soldier simulation a step closer to fielding.

Introduction

Future wars are likely to be fought in complex urban environments that range from skyscraper jungles to huge shantytowns (U.S. Army Training and Doctrine Command (TRADOC), 2002). According to U.S. Army doctrine "Small unit effectiveness and empowered leadership are critical to the success of these operations. Close urban assault has significant dismounted character, requiring a robust infantry capability to engage and sustain the urban fight (US Army TRADOC, 2002)."

Training to prepare for urban operations, also called military operations in urban terrain (MOUT), is currently limited to live exercises in small, not very complex MOUT villages. During the recent conflict in Southwest Asia, soldiers also trained in a Mobile MOUT facility consisting of prefab containers. Live training is essential for urban operations. However, MOUT villages have a number of limitations that constrain their effectiveness. Use of explosives and weapons is restricted by safety concerns and the cost of reconstruction. MOUT sites have proved difficult to instrument effectively limiting the information available for After Action Review. Soldiers quickly learn the layout of the MOUT village detracting from the realism of training in unfamiliar terrain. .Virtual environments could complement live environments to provide a MOUT training alternative whose strengths overcome live environments limitations.

Virtual environments have proven to be effective for training the U.S. Army's mounted forces

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

(Boldovici & Bessemer, 1994; Mastaglio, 2003). With the development of SIMNET and later the Close Combat Tactical Trainer (CCTT), armor and mechanized infantry units have been able to train on large terrain databases that are either geo-specific or geo-typical. They can employ weapons as they would be on the battlefield without the safety constraints of live ranges, and After Action Review (AAR) systems can record all the relevant activities that form the basis of discussion during AARs. In contrast, dismounted combatants have not played a major role in the Army's virtual training systems. The CCTT Dismounted Infantry Manned Module (DIMM) has been judged to be unacceptable by infantry soldiers. Early on there was not a lot of interest in the light Infantry community in the use of virtual simulations. However, this is changing. Recent conflicts in Afghanistan and Iraq have highlighted the need for alternative training methods for training dismounted combatants in MOUT (Third Infantry Division, 2003). The Commandant of the Infantry Center and School supports the development of a virtual dismounted combatant training capability. The operational requirements document for a Soldier Combined Arms Tactical Trainer or Soldier CATT is being staffed.

Soldier CATT will require a different approach and development of a new set of technologies to support simulation for dismounted combatants. CCTT trains soldiers in networked vehicle simulators. SIMNET, the predecessor to CCTT, applied innovative networking technology to simulation, but the simulators themselves use the same technologies previously used in aviation simulation (Thorpe, 1987). SIMNET and CCTT simulators place the crew in workstations which mirror those in actual combat vehicles. Visual and auditory displays present an environment that represents the virtual battlefield. Vehicle crews perform on this battlefield using the same controls they would use in the real world. Since dismounted combatants are not in vehicles, a virtual simulation capability must provide them with the means to operate in a virtual environment without the benefit of a traditional simulator. In a Soldier CATT the dismounted combatant must be provided with technologies that allow them to move, shoot and communicate in much the same way that they would in the real world. This presents a set of technical and practical obstacles that must be overcome to develop an effective Soldier CATT. Fortunately, Soldier CATT will have the benefit of Virtual Reality research that the U.S. Army has been conducting over the last ten years (Knerr et. al., 1998).

The Army's dismounted soldier simulation research has culminated in a four-year effort that ended in October 2002. The US Army Research Institute (ARI), US Army Simulation, Training and Instrumentation Command (STRICOM), and the US Army Research Laboratory (ARL)[1] worked together to advance the capabilities of dismounted soldier simulation technologies. The goal was to provide a virtual means by which infantry leaders and soldiers could receive meaningful MOUT or other combined arms training. Each of these US Army agencies had been working independently on problems associated with dismounted soldier simulation. That work formed the baseline from which the cooperative project started. This paper describes progress made in three technology areas key to dismounted soldier simulation and how the growth in these technologies impacted on soldier's impressions and performance. We compared soldiers' impressions at yearly user assessments from the first year of the research to its final culminating event at the end of year four. The three-technology areas are: representing the dismounted combatants' environment; producing realistically performing Dismounted Infantry Semi-Automated Forces; and developing of an After Action Review system that captures performance in MOUT.

The Dismounted Combatant Virtual Environment

[1] The elements of STRICOM and ARL conducting this research are now part of the US Army Research Development and Engineering Command (Provisional) (RDECOM)

The Squad Synthetic Environment (SSE) served as the baseline virtual environment throughout the course of the dismounted soldier simulation project. The SSE was produced as an outgrowth of earlier research conducted by STRICOM under the Dismounted Warrior Network program (Lockheed Martin, 1998). The SSE as it existed in 1998 was purchased by the U.S. Army Training and Doctrine Command (TRADOC) for use at the Dismounted Battlespace Battle Lab's Simulation Center. In the original purchase enough simulators were purchased to support a nine-man infantry squad. The SSE consisted of individual simulators known as Soldier Visualization Systems (SVS), produced by Reality by Design. The SVS is a personal computer-based, DIS compatible, dismounted infantry simulator that places soldiers in the middle of a square enclosure (approximately eight feet on each side). One side is a rear projections display on which the view of the virtual world is presented in 1024x768 resolution. The other three sides are sound-attenuation fabric. A surrogate weapon with an integrated thumb transducer is used for movement through the virtual environment. Posture changes (standing, kneeling, or prone) and weapon aiming are captured via an Intersense Corporation position tracking system. At the beginning of the four-year research effort the SVS was capable of displaying terrain data bases in daylight, without shadows

Figure 1. Soldier Visualization Station

or any other indication of time of day. Figure 1 shows a soldier in an SVS. The only MOUT terrain data base available was a model of the McKenna MOUT Village at Ft. Benning, GA. This data base was limited in that dismounted infantry semi-automated forces (DI SAF) could only enter one of the buildings. That building was the only one that was modeled so that SAF could navigate inside it. Scenarios that could be played out in the virtual environment of 1999 were limited to daylight attacks that emphasized movement techniques. Soldiers did not possess their full range of weapons (grenades, smoke). Buildings had to be entered through doors. It was a limited environment that left room for improvements.

Over the course of the research effort the dismounted combatant virtual environment improved dramatically. These improvements fit into three categories: lighting conditions, weapons and weapons effects, and dismounted soldier-oriented databases.

Updates to the SVS's ability to display lighting conditions included: streetlights, internal building lights, transformers, flashlights, night vision devices, visible and non-visible laser aiming lights, and shadows that change with time of day or night (Knerr et. al., 2002, 2003). Modifications were made to existing 3D Open Flight terrain data bases to include scene illumination from street lights and to place lights in buildings. The ability for lights to be "shot out" or turned off by blowing up a transformer was incorporated to improve movement at night by dismounted soldiers equipped with night vision devices. Immersed soldiers had the ability to use visible and non-visible aiming lights. Visible aiming lights could be seen by soldiers in SVS simulators and by DISAF. Non-visible aiming lights are only visible to those using SVS systems that have image intensification or night vision goggles. Building shadows were introduced to provide soldiers a place to hide. The shape of the shadows changes over to time as they would as the sun moves overhead.

The baseline SVS at the start of the research provided limited capabilities to influence events on the dismounted battlefield. Soldiers had individual weapons but they did not have available the other assets necessary to be successful in MOUT. During the last year of the research project attention was placed on adding the tools needed to succeed on the MOUT battlefield. Soldier reaction to initial user evaluations prompted the development and introduction of grenades (both fragmentation and concussion), flares (both airborne and thrown), tracer ammunition, and tactical smoke to cover movements. Satchel charges were also introduced in conjunction with dynamic terrain, allowing holes to be blown in the sides of buildings to facilitate entry. A Dynamic Terrain Server (DTServer) developed by ARL provides a means to blow holes in buildings sized appropriately for the munition and building material and to create rubble in addition to the hole. The DTServer transmits two types of results to receiving networked simulators. The first is a 'ding' packet, which results from small arms fire on a hard surface. Simulators receiving a 'ding' packet display a model of a simulated small crater at the point of impact. The second result is a 'breach'. 'Breaches" could be caused by satchel charges or anti-tank rounds. Receiving simulators of a 'breach' packet would replace polygons in the explosion area with new ones that could be transited by soldiers. Development of the DTServer provided the means to compute and distribute dynamic terrain changes to the simulation network (Thomas, 2003).

Two terrain data bases were developed in addition to the McKenna MOUT Village. One represented the Shuggart-Gordon MOUT Village at Ft. Polk, LA. The other was a notional terrain data base added onto the original McKenna MOUT site. The additions included two new high rise structures (one twenty stories and the other ten stories), numerous single-story buildings, a bridge, and a set of tunnels running below the original McKenna part of the data base. Multiple Elevation Structures were built into all of the buildings in each of the data bases. This allowed DISAF entities to enter and operate in each of the buildings and on all of the floors.

Improvements to DISAF

DISAF was developed to provide a realistic representation of dismounted infantry and civilians on the virtual battlefield. The primary focus of DISAF has been the development of tactical behaviors for individual through squad level operations. Our original aim was to focus on training small unit leaders and DISAF would be used to fill in squads or fire teams to support training of leaders. DISAF would also provide a capable enemy, and fill out the battlefield with armed and unarmed civilians.

DISAF is based on the OneSAF Test Bed (OTB) architecture. Most of the DISAF behaviors are based on validated Combat Instruction Sets. DISAF runs on a PC under Linux or Windows NT. A DISAF operator uses an execution matrix to preplan the movements and behaviors of DISAF. In the first year of the research we found that it was difficult for the DISAF operator to change where a DISAF entity was going or what he was doing during the course of the exercise. This delay in DISAF reacting to new instructions was unacceptable to soldiers. A major effort was put into producing a more effective DISAF operator interface and developing a capability for small unit leaders to direct DISAF through voice commands. The voice control of DISAF was to provide a means for live and virtual participants in a simulation to interact through spoken statements. The live user can maintain command and control over synthetic entities (DISAF) while the synthetic entities can vocally acknowledge command and provide information to the live participant.

Development of DISAF began under STRICOM's Dismounted Warrior Network project. Over the course of this four year effort behaviors and capabilities were added on a yearly basis. DISAF are organized as individuals and units both enemy, friendly, and civilian. They possess behaviors that allow them to operate in the open and in buildings. During the final year of the project behaviors were added for armed civilians, crowds of civilians, wounding of entities to include visual signs of bleeding. Knerr et. al. (2003) contains a complete list of entities, behaviors and capabilities for DISAF. DISAF has been included as a component of OneSAF Test Bed.

Dismounted Infantry After Action Review System (DIVAARS)

DIVAARS was developed to meet two needs. The first was to provide soldiers with a common understanding of what happened during an exercise and why it happened, so that they can identify ways to improve their performance. The second was to facilitate data analysis, in order to support both training feedback and research and development. Determining what happened during an exercise is particularly difficult in an urban environment, because buildings and other structures break up the visual field and limit the portion of the battlefield that can be observed by any one person.

The AAR system connects to the network used by the soldier simulators and DISAF, and permits observation and recording of the exercise data. AAR Leaders are central to the design and operation of DIVAARS. They observe what happens during the conduct of an exercise and prepare a presentation that will lead the unit to an understanding of what happened, why it happened, and how to do better. Their presentation should be both interactive and efficient.

A description of key DIVAARS capabilities follows. The emphasis is on unique DIVAARS capabilities. Figure 3 shows a sample DIVAARS display with many of these features.

Playback. Playback controls include actions such as pause, stop, record, play, step forward, fast-

Figure 2. AAR system main display.

forward, rewind, fast reverse, and step reverse. Variable playback speeds are available. The AAR

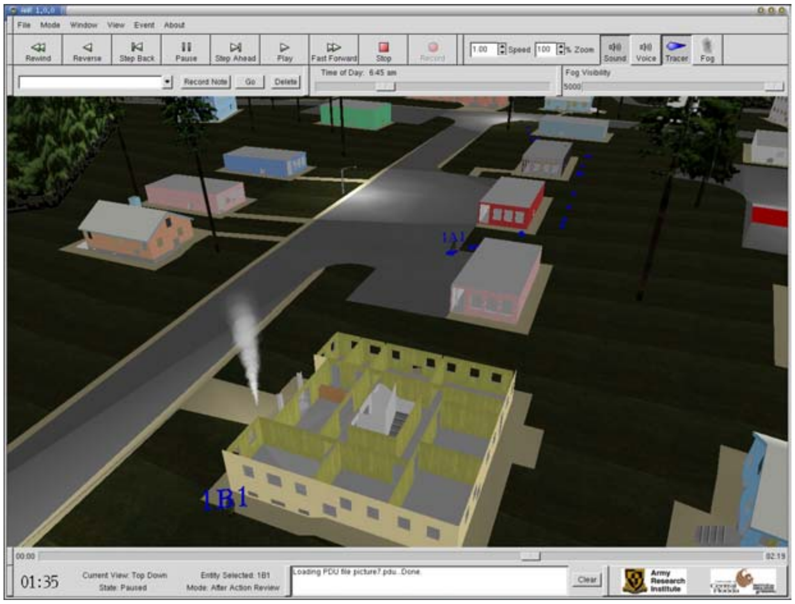
Leader also has the capability to mark events during the exercise, and jump directly to them during the AAR.

Viewing Modes. Multiple viewing modes are available during both the exercise and the AAR.

- Preset Views – An unlimited number of preset views can be selected at any time prior to or during the exercise for immediate use.

- Top-Down – A view of the database looking straight down from above. It can be moved left, right, up, down, and zoomed in or out. It can also be locked onto an entity, in which case it will stay centered

Figure 2. DIVAARS Screen and Controls



directly above that entity as it moves through the database.

- 2D View - This is the traditional plan view display. It is the same as Top-Down except that depth perspective is not shown.
- Entity View – Displays what a selected entity (including enemy or civilian) sees, including the effects of head turning and posture changes.
- Fly Mode – The AAR Leader can "fly" through the database using the mouse for control.

Movement Tracks. Movement tracks show, in a single view, the path an entity travels during an exercise. Markers are displayed at fixed time intervals.

Entity Identifier. Friendly force avatars in the DIVAARS, as in the virtual simulators, are identical. A unique identifier, as given in the entity marking field of the DIS Entity State Protocol Data Unit (PDU), is shown above the avatar of each unit member.

Digital Recording and Playback of Audio Program. DIVAARS records and plays back audio content for all scenarios.

Viewing Action inside a Building. The AAR Leader can select a building and then select a floor of that building to be displayed. Using this feature, the operator can view and display the avatars going through a building without the problem of upper floors or outer walls blocking the view.

Dynamic Terrain Changes. DIVAARS receives the PDU messages from the DTServer and updates the display with any changes.

Bullet Lines. Bullet flight lines are shown for all weapon firings. The line traces a shot's origin and destination. It is the same color as the originating entity. These bullet lines gradually fade away.

Event Data Collection and Display. DIVAARS has the capability to track many events including shots fired, kills by entities, movement, and posture changes. These data can be shown in a tabular format or graphical display. Ten different tables and graphs are available:

- Shots fired, by entity and unit
- Kills, by entity and unit
- Killer-Victim table that shows who killed whom, the angle of the killing shot (front, flank, or back), and the posture of the victim (standing, kneeling, or prone)
- Shots as a function of time, by entity, unit, and weapon
- Kills as a function of time, by entity, unit, and weapon
- Kills by distance from killer to victim, by entity, unit, and weapon
- Rate of movement of each entity, and averaged at team/squad levels
- Percentage of time friendly units were stationary
- Percentage of time friendly units were in different postures
- Display of user-defined events

User Evaluations

Methods

At the end of each year of the research and development effort, Culminating Events (CEs) were held to insure the compatibility of individual technologies under development, assess their usability in a realistic setting, and obtain soldier feedback on their use and effectiveness. While the specific details of the CEs varied from year to year, those for the first, third, and fourth years were very similar. Each involved the conduct of a series of tactical urban scenarios by full or partial squads of Infantry soldiers in networked immersive simulators, the SSE. DISAF served as enemy and civilians, and filled some friendly positions within the squads.

Soldiers reported to the Dismounted BattleSpace BattleLab Virtual Simulation Lab, Fort Benning, Georgia, for either a one- or two-day period. Upon their arrival, they were briefed on the purpose of and procedures for the exercises, assigned duty positions, and given instruction and practice on the use of the simulators. A series of tactical exercises sessions followed. Each session consisted of delivery of the mission order, squad leader development of the mission plan and brief to his squad, conduct of the mission, and an AAR. After the AAR questionnaires were administered and interviews conducted.

We were interested in the how well the technology permitted the soldiers to perform their required tasks, their perceptions of the individual technologies, and how much they learned during their training. We made extensive use of questionnaires and interviews. While there were variations in some of the questionnaire items from year to year, there was a common set of questions used in all CEs. The *Simulator Capability Questionnaire* asked soldiers to rate their ability to perform various tasks in the simulators. The *SAF Performance Questionnaire* asked leaders to compare the performance of SAF with that of real soldiers. The AAR questionnaire asked about the effectiveness of the AAR. The *Training Effectiveness Questionnaire* asked leaders only how much they thought their performance improved during the exercises.

The simulator network configuration for the Year 4 CE is shown in Figure 3. The following items were connected to the network:

- Six SVS individual soldier simulators used by the squad leader, two fire team leaders, and three Fire Team A members. All SVSs were equipped with headsets which permitted verbal communication.
- One Voice Recognition PC
- Two DIVAARS Systems
- One Dynamic Terrain Server
- One BattleMaster/DISAF Operator Station. The DISAF Operator and the Exercise Controller used this station.
- One Desktop SVS used by a role player

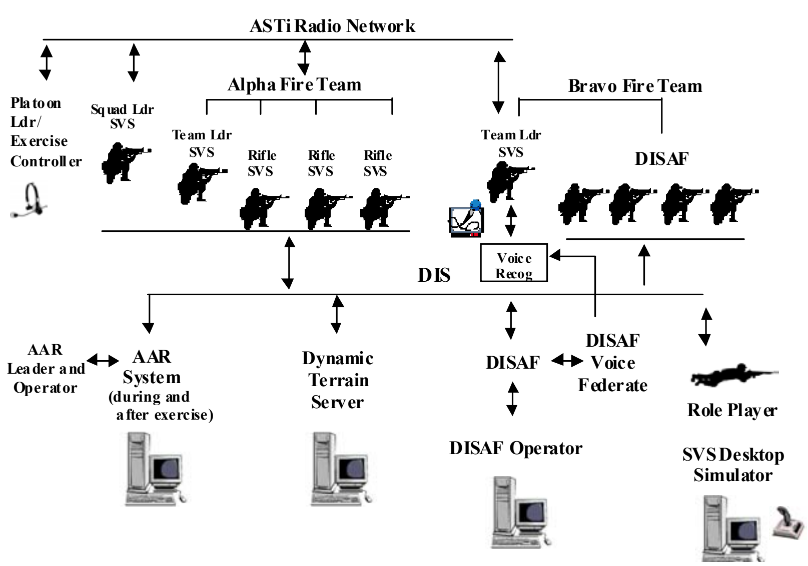


Figure 3 Year 4 CE Configuration

All soldiers completed the Simulator Capability Questionnaire. Scores were calculated for each task by assigning a response of *Very Poor* a value of 0, *Poor* a value of 1, *Good* a

value of 2, and *Very Good* a

value of 3. Table 1 provides a comparison of the simulator capability ratings across Years 1, 3, and 4. The items are listed in order of descending Year 4 mean value. The most noticeable result is a fairly consistent pattern of higher ratings in Year 4 than in Year 3 and, to a lesser extent, Year 1. The overall mean of the common items was 1.90 in Year 1, 1.74 in Year 3 and 2.12 in Year 4.

Thirty-six of 52 tasks were rated *Good* or higher (mean equal to or greater than 2.0) in Year 4, as compared with 16 in Year 1 and 14 in Year 3. The more highly rated tasks consisted of identification of types of people (such as civilians and non-combatants) and tactically significant areas, imprecise movement, and communication. The lower rated tasks consisted of precise or rapid movement, including aiming, distance estimation, and locating the source of enemy fire using either visual or auditory cues. Twenty-two Year 4 item means were significantly higher than the same items in Year 3, and 20 were higher than corresponding Year 1 means (p<.05). Items for which the means changed significantly are shown in Table 1.

Table 1

Simulator Capability Questionnaire Responses

Task	VE STO Year 1 Mean[a]	VE STO Year 3		VE STO Year 4	
		Mean	N	Mean	N
Execute planned route.	1.89*	2.06*	18	2.67	18
Identify assigned sectors of observation.	2.06*	1.94*	17	2.53	17
Move in single file.	2.00*	1.94*	18	2.50	18
Look around corners.	1.47*	1.29*	17	2.50	18
Communicate enemy location to team member.	2.06*	1.89*	18	2.50	18
Understand verbal commands.	1.94*	2.29	18	2.47	18
Fire weapon in short bursts.	2.00*	1.89*	18	2.44	18
Move quickly to the point of attack.	1.94*	1.89*	18	2.44	18
Communicate spot reports to squad leader.	1.94*	2.00*	18	2.44	18
Scan from side to side.	1.72*	1.94*	18	2.44	18
Identify sector of responsibility.	2.11*	2.17	17	2.39	18
Identify civilians.	2.72*	2.22	18	2.33	18
Coordinate with other squad members.	1.88*	2.00	18	2.33	18
Execute the assault as planned.	1.89*	1.83*	18	2.33	18
Locate support team positions.	2.00*	1.72*	18	2.33	18
Identify covered and concealed routes.	1.94*	1.94*	18	2.28	18
Identify safe and danger areas.	2.22	2.11*	18	2.28	18
Maneuver below windows.	2.06	1.61*	18	2.22	18
Locate buddy team firing positions.	1.94	1.78*	18	2.22	18

Engage targets within a room.	2.06	1.61*	18	2.22	18
Identify enemy soldiers.	2.44	1.53*	17	2.22	18
Identify areas that mask supporting fires.	1.72*	2.00	18	2.17	18
Take hasty defensive positions.	1.89	1.71*	17	2.11	18
Determine other team members' positions.	1.78*	2.00	18	2.06	17
Scan the room quickly for hostile combatants.	1.76	1.29*	17	2.06	17
Maintain position relative to other team members.	1.78*	2.06	18	2.06	18
Maneuver around corners.	1.67*	1.06*	18	2.00	18
Maneuver around obstacles.	1.67*	1.39*	18	1.94	18
Estimate distances from self to a distant object.	1.72	1.22*	18	1.89	18

Notes: Year 1 N=18. A blank in a cell indicates that that question was not included in that year.

*** significantly different from the Year 4 mean at p<.05.*AAR System***

All soldiers rated the AAR system on eight items. Table 2 shows the percentage of the soldiers who agreed or strongly agreed with those positively-worded statements about the effectiveness of DIVAARS in both Year 3 and Year 4. The ratings likely reflect the combined performance of the AAR Leader and DIVAARS. DIVAARS was a tool the AAR Leader used to analyze problems with the unit execution, determine the causes of problems, and then facilitate participant dialogue. The ratings are overall very high, with at least 94% of the soldiers agreeing or strongly agreeing with every item.

Table 2
AAR System Ratings

The AAR system	Rating	Year 3 (N=18)	Year 4 (N=17)
...was effective in displaying movement outside of buildings	***Strongly Agree*** Agree Total	28% 61% 89%	82% 18% 100%
...was effective in displaying movement inside of buildings	***Strongly Agree*** Agree Total	33% 50% 83%	82% 18% 100%
...was effective in replaying communications	***Strongly Agree*** Agree Total	28% 28% 56%	82% 12% 94%
...made clear what happened during a mission	***Strongly Agree*** Agree Total	44% 56% 100%	82% 12% 94%
...made clear why things happened the way they did during a mission	***Strongly Agree*** Agree Total	44% 39% 83%	76% 24% 100%
...made clear how to do better in accomplishing the mission	***Strongly Agree*** Agree Total	28% 56% 84%	71% 24% 95%
...made clear the order in which key events occurred during the mission	***Strongly Agree*** Agree Total	33% 67% 100%	82% 12% 94%
...was more effective than conducting an AAR without any visual or auditory playback (just talking)	***Strongly Agree*** Agree Total	50% 33% 83%	94% 6% 100%

DISAF Performance

The Squad Leaders and Fire Team Leaders rated DISAF performance. Results are shown in Table 3. A rating of 0 indicated that the DISAF were about the same as real soldiers, while a +1 indicated they were slightly better, and –1, slightly worse. Ratings in Year 4 improved relative to Year 1 and Year 3. DISAF can locate/identify the enemy better than real soldiers, but have trouble moving to and firing at the correct locations, and reporting their observations or activities to their Fire Team Leader. Generally, those activities rated lower in Year 4 than in Year 1 were activities, primarily control of movement, that were performed by

the DISAF operator in Year 1 and by an immersed leader via voice control in Year 4. Thus while the introduction of voice control may have reduced the workload of the DISAF operator, it was not necessarily an improvement from the leader's perspective.

Table 3

Mean DISAF Behavior Ratings

SAF Behavior	Year 1 (N=9)	Year 3	Year 4 (N=9)
Distinguish between friendly and enemy positions.	-0.90	0.22	0.89
Locate known or suspected enemy positions.	-1.10	-0.22	0.67
Clear a room.	-1.44	-0.44	0.56
Fire weapons accurately.	-0.40	-0.43	0.22
Clear a building.	-1.33	-0.38	0.11
React to contact.	-1.00	-0.89	0.00
React to ambush.	-0.88	-1.11	-0.22
Move through open areas.	-1.00	-1.00	-0.67
Take hasty defensive positions.	-0.50	-1.11	-0.67
Maintain position relative to other squad or team members.	-0.50	-1.29	-0.67
Deliver suppressive fire.	-0.80	-0.88	-0.78
Support by fire.	-0.67	-1.38	-0.78
Move through built-up areas.	-0.67	-1.00	-0.89
Move to designated location.	-0.10	-0.13	-1.11
Perform fire and movement.	-0.67	-1.00	-1.25
Communicate information to squad leader.	-0.80	-1.11	-1.38
Change formation.	-0.62	-1.25	-1.56
Mean	**-0.79**	**-0.79**	**-0.44**

Note. Year 3 N varies from 7 to 9.

Training Effectiveness

Generally, Squad and Fire Team Leaders said that their performance improved as a result of the training. The percentage who said that their performance improved at least slightly ranged from 82% for the task "Clear a building" to 100% for "Assess the tactical situation," "Control your squad or fire team," and "Plan a tactical operation." Year 4 ratings were generally better, and in only one case worse, than those given to the same tasks in Year 3 CE. Complete results are shown in Table 4. In general, ratings for coordination, communication, and control tasks were higher than those for specific unit tasks or battle drills, although this difference was not as pronounced in Year 4 as it had been in previous years.

Table 4
Squad and Fire Team Leader Training Effectiveness Ratings

Task	% Indicating Improvement		
	Year 1	Year 3	Year 4
N	9	15	18
Assess the tactical situation.	67%	93%	100%
Control your squad or fire team.	67%	80%	100%
Plan a tactical operation.	33%	73%	100%
Squad/fire team communication and coordination.	78%	80%	94%
Control squad or fire team movement during assault.	67%	80%	89%
React to Contact Battle Drill.	44%	80%	89%
Locate known or suspected enemy positions.	44%	67%	89%
Coordinate activities with your chain of command.	44%	100%	88%
Control squad or fire team movement while not in contact with the enemy.	67%	80%	83%
Clear a room.	44%	53%	83%
Clear a building.	56%	57%	82%[a]

Note. Squad and Fire Team Leaders who participated for two days completed the questionnaire at the end of each day.

[a]N = 17

Discussion

Perhaps the most significant accomplishments of the VE STO are not reflected in the ratings or performance data that were collected but in the level of sophistication and complexity of the scenarios that were run. In the Year 1 CE, at the end of the first year of the STO, five different scenarios were used. All were basically the same: initiate movement to an objective building, react to enemy contact in route, resume movement and finally assault the

building. It was always daylight. DISAF could not enter buildings. Few civilians were present, and their behaviors were limited to either standing still or moving on a preplanned route. Buildings could not be breached. Neither force could use smoke or grenades. A hit always equaled a kill. A fire team leader could control DISAF only by giving a verbal command to the DISAF operator, who then implemented that command at his console. Routes for DISAF had to largely be scripted in advance. AARs were limited to linear playback on a stealth viewer. In Year 4, there were six different scenarios. Scenarios could be conducted at any time of day or night. DISAF could go anywhere, and could carry out some highly sophisticated behaviors autonomously, such as room clearing. Civilians moved about freely, as individuals and in crowds, and could be armed. Holes could be blown at any location in any building. Flares, smoke, and grenades were available to all participants. Soldiers could be wounded as well as killed when hit. These factors greatly increased the variety and realism of the training situations that could be presented.

Simulator Capabilities

While it was satisfying to find that soldier ratings of the simulator capabilities were generally higher than in previous years, it was difficult to relate the changes in rating on specific items to a likely cause. For example, why did soldiers give the task "move in single file" a higher rating in Year 4 than in Year 3? While capabilities have been added to the SVSs, the basic characteristics remain the same. The most likely explanation is that the soldiers responded to the individual items on the basis of both the specific item content and their perception of the overall quality of their experience in the simulators. The new capabilities, like smoke and grenades, which were rated highly (and the absence of which was a cause for complaint in prior years), may have increased the overall quality of this experience and, by extension, the ratings of individual tasks that were not directly affected.

Other factors may have had less straightforward but nevertheless substantial effects on the ratings. As described above, the training scenarios have become increasingly challenging and complicated over the course of the STO. While this made the training more realistic, it required the soldiers to try to perform more complicated tasks in the simulators, and may also have made it more likely that the soldiers would encounter the limits of the simulators. The video gaming experience of the soldiers may also have been a factor in the ratings. It appeared from the interviews and informal interactions that the game-playing experience of the soldiers has increased over the years. On the one hand, their gaming experience has given them opportunity to acquire necessary "basic skills" that make it easier to learn to function in the SVSs. One group of soldiers reported in their interviews that they had no difficulty learning to use the SVSs because "We're the Nintendo generation." On the other hand, the impact of the increasing sophistication of computer and video games may have caused soldiers to have higher standards for simulator performance. The simulator capabilities are being compared with increasingly realistic and sophisticated commercial products. This has, in effect raised the standards by which automated entities and environments are judged.

Training Effectiveness

Leader ratings of training effectiveness constitute perhaps the biggest success story of the STO. Since Year 1, we have seen a consistent increase in leader ratings of training effectiveness. Like the ratings of simulator capability, these ratings were likely influenced by the changes in the backgrounds and experience of the leaders and administrative changes (primarily the separation of the roles of the O/C and the AAR Leader).

There is a broad range of tactical skills that could conceivably be trained in VE. At one end of the continuum are small unit leader decision-making skills. Pleban, Eakin, Salter, and Matthews (2001) found that these skills could be trained effectively in VE. Training these skills does not require a high fidelity, fast, or precise interface with the virtual world. Success is more likely to depend on the scenarios and the quality of the role-players. At the other end of the continuum are the specific squad drills and tasks, like building clearing, which involve less decision making, more communication and coordination among unit members, but above all require rapid and precise positioning, movement and use of weapons. A recent experiment by Pleban and Salvetti (2003) indicates that, while there are a number of interface and technology problems to be overcome, VE nevertheless shows promise for this type of training as well, although it appears not to be effective as real world training at present. The types of squad-level exercises conducted during the last two CE's fall somewhere in the middle, targeted at improving leader decision-making and command and control skills in a variety of mission types.

Given the current state of technology, it does not appear that VE is an effective complete replacement for real world tactical training. However, it could be used effectively for some types of training and some stages of training. VE training could provide the walk phase of the training, concentrating on improving the decision-making, situation awareness, communication, and coordination skills. Real world training would place greater emphasis on the motor skills. VE training also has the advantage of being more flexible, in that terrain databases and environmental conditions can be changed more rapidly than a real world urban training center.

AAR System

DIVAARS performed very well, although in looking at the DIVAARS ratings it must be kept in mind that the AAR is a product of the combination of a skilled AAR Leader and the AAR system.

DISAF Performance

DISAF capabilities have increased enormously since the start of the VE STO. This permits more realistic scenarios. DISAF does some things better than others. For example, they are quite good (better than human soldiers) at detecting and firing upon the enemy, but control of their movement, particularly in a dynamic situation, is awkward.

Conclusions and Future Directions

Substantial improvements have been made during the last four years in the capability of virtual simulation to provide training for the leaders of small dismounted Infantry units. These developments in technology have greatly increased the level of realism that is possible through virtual simulation, and the breadth of tasks that can be trained. While the samples are small, both leader self-ratings and independently-obtained performance scores during this CE indicate that soldier skills improved with practice in VE. Moreover, leader-self-ratings of skill improvement have increased regularly since the first year of the STO. The Year 3 and Year 4 CEs have focused on sustainment and support operations, and in that context, the leaders reported more improvement in command and control, coordination and communication, planning, and situational awareness skills than in skills conducting specific unit tasks or battle drills. Similarly, Pleban et al. (2001) found VE effective for training platoon leader decision making skills.

Given the current state of technology, it appears that VE could be used effectively for some types of training and some stages of training. VE could be used for the walk phase of the training, concentrating on improving the decision making, situation awareness, communication, and coordination skills, while real world training could place greater emphasis on the motor skills. The level of interest in virtual simulation has increased in the Infantry community as evidenced by the current interest in Soldier CATT. This research effort has made significant progress in developing a capable training system. The next step should be an advanced development effort, taking a total systems approach, to produce a prototype VE training system for the dismounted Infantry leaders and soldiers training for MOUT.

References

Boldovici, J.A. and Bessemer, D.W. (1994) *Training Research with Distributed Interactive Simulations: Lessons Learned from Simulation Networking* (Technical Report 1006). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knerr, B.W., Lampton, D.R., Singer, M.J., Witmer, B.G., & Goldberg, S.L. (1998*) Virtual Environments for dismounted soldier training and performance: Results, recommendations, and issues* (ARI Technical Report 1089). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knerr, B.W., Lampton, D.R., Crowell III, H.P., Thomas, M.A., Comer, B.D., Grosse, J.R., Centric, J.H. Garfield, K.A., Martin, G.A., & Washburn, D.A. (2002) *Virtual Environments for Dismounted Soldier Simulation, Training and Mission Rehearsal: Results of the 2001 Culminating Event* (Technical Report 1129). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knerr, B.W., Lampton, D.R., Goldberg, S.L., Thomas, M.A., Comer, B.D., Grosse, J.R., Centric, J.H., Blankenbeckler, P., Dlubac, M., Wampler, R.L., Siddon, D., Garfield, K.A., Martin, G.A., & Washburn, D.A. (In Press, 2003) *Virtual Environments for Dismounted Soldier Simulation, Training and Mission Rehearsal: Results of the 2002 Culminating Event* (Technical Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lockheed Martin Corporation (1998). *Dismounted warrior network enhancements for restricted terrain* (Advanced Distributed Simulation Technology II, Dismounted Warrior Network DO #0055). Orlando, FL: U.S. Army Simulation, Training and Instrumentation Command.

Mastaglio, T. (2003) *Assessing the Effectiveness of the Close Combat Tactical Trainer* (Final Contract Report) Portsmouth, VA: MYMIC Corp.

Pleban, R. J., Eakin, D. E., & Salter, M. S., & Matthews, M. D. (2001). *Training and assessment of decision-making skills in virtual environments* (Research Report No. 1767). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA389677)

Pleban, R. J., & Salvetti, J. D. (2003). *Using virtual environments for conducting small unit dismounted mission rehearsals* (Research Report No. 1806). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Third Infantry Division (2003) *Operation Iraqi Freedom* (Draft After Action Review). Ft. Stewart, GA: Third Infantry Division.

Thomas, M.A. (2003) *The U.S. Army Research Laboratory Dynamic Terrain Server* (ARL-TR-2962). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Thorpe, J.A. (1987). *The new technology of large scale simulator networking: Implications for mastering the art of warfighting.* In the Proceedings of the 9th Interservice Industry Training System Conference (pp492-501), Washington, D.C.: American Defense Preparedness Association, November, 1987.

U.S. Army Training and Doctrine Command. (July, 2002) *Objective Force Maneuver Units of Action* (U.S. Army TRADOC Pamphlet 525-3-90). Ft. Monroe, VA: Headquarters, U.S. Army Training and Doctrine Command.

Virtual Technologies and Environments for Expeditionary Warfare Training

LT Joseph Cohn, PhD
US Naval Research Laboratory
Code 5583
4555 Overlook Ave., SW
Washington, DC 20375-5337

LCDR Dylan Schmorrow, PhD
DARPA/IPTO
3701 Fairfax Drive
Arlington, VA 22203-1714

Dr. Denise Lyons
Naval Air Warfare Training Services Division
NAVAIR Air 4962: Training Technology
12350 Research Parkway
Orlando, FL 32826-3224

Dr. James Templeman
US Naval Research Laboratory
Code 5511
4555 Overlook Ave., SW
Washington, DC 20375-5337

Peter Muller
Visitech, Inc.
1221 Shaker Dr.
Herndon, VA 20170

Summary

Over the past decade, Virtual Environment (VE)-based training systems have become commonplace within the military training domain. These systems offer such benefits as small footprint, rapid reconfiguration, and enhanced training delivery. In addition, they appear to offer significant relief for a market starved for low cost training systems, and hold great potential as effective training tools. Yet, all too often the human element is taken for granted, with systems being designed to incorporate the latest technological advances, rather than focusing on enhancing the user's experience within the VE -both from a training and human factors perspective. It is precisely this shift in design philosophy, from techno centric to human centric that represents the next, greatest, challenge to developing effective VE-based training systems.

Interaction with VE involves the ability of individuals to effectively perform essential perceptual-sensory-motor tasks within the virtual world. More specifically, this can involve the ability to move about the VE, manipulate virtual objects, locate virtual sounds, deal appropriately with physical constraints, or perform visual tasks (i.e., discriminate colors; judge distance; search for, recognize, and estimate the size of objects). Interactive technologies include multi-modal 3D displays and input devices, real-time rendering, and distributed simulation (i.e., multiple user interaction through networked VE systems). These technologies define how the environment is portrayed and how it responds to user actions. The design, synthesis, and analysis of new interaction technologies will be based on our growing understanding of human perception and action in VEs. Tools are needed to provide a more comprehensive assessment of the quality of interaction.

The Office of Naval Research's Virtual Technologies and Environments (VIRTE) Program was developed to address this design challenge. VIRTE is focusing on developing two capabilities central to applying VE technology to training: (1) improving the quality of interaction provided by VE, and (2) applying advanced training aids and methodologies to real Navy and Marine Corps requirements. Quality interaction is essential for making VE usable for extended training and for applying VE

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

technology to build combat simulators. Advanced training aids and methodologies are being developed to derive full benefit from the flexible new medium of VE.

Our approach takes maximum advantage of the growing knowledge base in human-centric design to build systems that are well-suited for enhancing performance of real-world tasks. In addition to traditional Knowledge Acquisition and Knowledge Engineering approaches, our interdisciplinary development team is utilizing team task analyses, motion sickness evaluations, and usability analyses. The team has also developed a novel methodology for evaluating team-based performance enhancement in VE systems. This cross-disciplinary approach gives the team a unique perspective into designing effective, low cost, deployable simulations. The technology testbeds that are being developed are useful not only for stand-alone training, but also for embedded training, team training, and mission rehearsal.

Introduction

Virtual Environment training Systems

Throughout the past decade, the United States military training community has turned its focus to Virtual Environment (VE) technologies as a means for providing rapidly deployable, easily configurable, affordable training solutions. As a result, a veritable barrage of (VE)-based training systems have been developed as potential solutions to a range of training requirements. While these systems typically offer the latest, cutting edge visual and hardware technologies, all too often the human element is taken for granted (Cohn, Breaux, Nguyen Schmorrow, 2001). A direct result of this techno-centric design philosophy is that the effectiveness of these systems is all too-rarely objectively evaluated. Consequently, the likelihood of designing a final product with a strong ‘gee whiz’ factor, combined with, at best, an undefined ‘training effectiveness’ quotient, is quite high, leading users to over-estimate the value of their training. In the operational setting, these over-estimations can have costly –if not tragic- consequences.

Human-centric Design Principles

Yet this need not be so. In parallel to the quantum leaps made in the technology realm, human factors specialists have made significant strides in developing human-centric design methodologies for ensuring that training systems satisfy the needs of the users for whom they are intended. The crucial element that has been missing is an underlying philosophy linking these techniques in a logical fashion.

Real World Example

Virtual Technologies and Environments (VIRTE)

The positive impact resulting from implementing these principles is best illustrated through an example of an actual VE training program, the Office of Naval Research’s Virtual Technologies and Environments (VIRTE) Program. VIRTE spans several levels of simulation and will ultimately provide a networked, deployable system for training a myriad of concepts relating to a specific type of warfare known as Expeditionary Warfare. Expeditionary Maneuver Warfare (EMW) … “is the union of our core competencies; maneuver warfare philosophy; expeditionary heritage; and the concepts by which we organize, deploy, and employ forces” (Department of the Navy, 2001). Naval Expeditionary Maneuver Warfare consists of military operations mounted from the sea, usually on short notice. They are carried out by forward-deployed or rapidly deployable, self-sustaining naval forces tailored to achieve a clearly stated objective. The future primary platforms for Expeditionary Warfare, known as the “Amphibious Assault Triad”, are the Landing Craft, Air Cushion, or LCAC, the Advanced Amphibious Assault Vehicle (AAAV), and the Osprey MV-22 tilt-rotor aircraft. VIRTE focused on these diverse vehicles since they present unique simulation and training challenges, as well as a component contributed by the United States Marine Corps, supporting Close Quarters Battle.

Within the VIRTE program, a range of simulations are currently under development. Each of these systems is based on real-world operational requirements and each is intended to transition to real world use as shown in Figure 1.

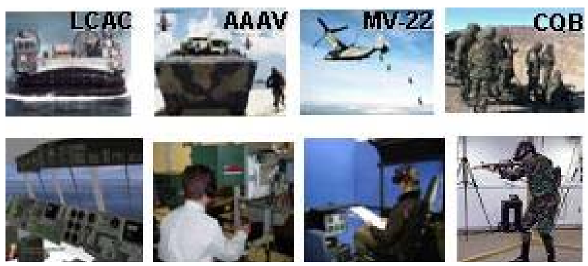


Figure 1: VIRTE focuses on providing deployed training solutions to the Expeditionary Warfare community. This community includes air, land and sea elements. *From Far Left to Right Top Row:* Landing Craft Air Cushioned (LCAC); Advanced Amphibious Assault Vehicle (AAAV) The MV-22 (Osprey) is a tilt-rotor craft currently under development; Close Quarters Battle. *From Far Left to Right, Bottom Row:* Virtual Environment (VE) LCAC; Virtual Environment (VE) AAAV; Virtual Environment (VE) MV-22; Virtual Environment (VE) CQB.

Landing Craft Air Cushioned (LCAC)

The LCAC is the only member of the triad that is currently fielded. The LCAC is a high-speed, over-the-beach fully amphibious landing craft capable of carrying a 60-75 ton payload. It is used to transport weapons systems, equipment, cargo and personnel from ship to shore and across the beach. The Navy has 76 LCACs in service. The LCAC crew consists of three positions, the Craftmaster (pilot), the Navigator, and the Engineer who work closely together to operate the vehicle. Currently, there are two LCAC Full Mission Trainers (FMTs), one at Assault Craft Unit (ACU) 4 in Dam Neck, Virginia, and one at ACU 5 in Camp Pendleton, California. While these provide excellent high fidelity training, they are expensive to procure and operate. The LCAC fleet is just beginning to field a Service Life Extension Program (SLEP) which completely changes the operator interface to the vehicle. It will be several years until there are sufficient SLEP LCACs fielded to transition the FMT to the SLEP configuration.

Advanced Amphibious Assault Vehicle (AAAV)

The Advanced Amphibious Assault Vehicle (AAAV) is currently in the prototype stage and it is scheduled to enter Low Rate Initial Production in FY 07. The AAAV will provide the capability to move a combat loaded USMC rifle squad at over 20 knots on the water and maneuver cross country with the speed and agility of the M1 tank. It will replace the AAV7A1. The AAAV crew consists of the driver, the gunner, and the vehicle commander. The AAAV is in the System Development and Demonstration (SDD) phase and second generation prototypes are being matured and prepared for production. The AAAV will be supported by a significant investment in training systems. VIRTE is focusing on transferring technology to two of these, the schoolhouse training system and the vehicle embedded training system.

MV-22

The MV-22 Osprey is a tilt-rotor aircraft that will provide airlift in support of Expeditionary Maneuver Warfare. The tilt-rotor design combines the speed, range, and fuel efficiency normally associated with turboprop aircraft with the vertical take-off/landing and hover capabilities of helicopters. The MV-22 crew consists of a pilot and co-pilot. The MV-22 is currently in Low Rate Initial Production (LRIP) and in the USMC will replace the CH-46E and CH-53D. As with the AAAV, there is a significant investment in training systems. Rather than try to insert technology directly in the MV-22 program, we are concentrating on demonstrating technologies that are applicable to all aircraft trainers.

Close Quarters Battle Military Operations in Urban Terrain (CQB for MOUT)

CQB for MOUT is a type of urban warfare that is distinct from the commonly portrayed ‘jungle’ or ‘desert’ combat scenarios. In CQB for MOUT, small teams, typically in groups of 4, are tasked with clearing a building of enemy units, room by room. It is a type of warfare that is extremely hazardous and man-power intensive. Currently, there only two options for providing CQB for MOUT training. The first option is Live training, which involves moving these teams through mock-up villages (‘combat towns’) using simulated ammunitions. The second option is through the use of simple computer generated scenarios, involving the projection of scenes on a screen, with users navigating through the scenario using simple interfaces such as joysticks or footpedals.

Both training options have significant disadvantages. Live training has significantly reduced realism due to the numerous safety constraints imposed during such activities. As well, this type of training involves the use of pre-made building mock-ups which can not be easily reconfigured, greatly reducing the range of potential training scenarios. At the same time, the simulations currently in use also fail to provide a strong basis for performance enhancement. Current systems utilize non-natural Human-Computer Interfaces and therefore cannot teach tactical mobility skills. Further, the manner in which the computer-generated scenes are slaved to the user’s movement does not support the development of the ‘move-look-shoot’ philosophy crucial to this form of combat. Finally, although CQB for MOUT is a team-based activity, current training systems, which present one scene to an entire team, can not support any level of ‘individualized’ first person perspectives that is part and parcel of this combat environment.

Prototype Development

While the real-world analogues of each VIRTE component serve unique roles in actual combat and impose unique training requirement upon their operators, their efforts complement each other; standard military doctrine relies on elements from each to achieve the overall objective (Department of the Navy, 1999). In parallel, the virtual systems are designed with the dual purpose of supporting training at the individual level, as well as at the level of distributed, team-based events, operating within a shared synthetic battlepsace.

In order to develop a suite of VE-based training tools that supported each of these Expeditionary Warfare components, within the Department of Defenses’ modeling and simulation framework, several parameters had to be defined. First, the technologies underlying each VE system must be compliant with the US Department of Defense (DOD) High Level Architecture (HLA). Second, each of the simulations had to share the same virtual battlespace. Finally, since these systems were intended to transition directly to the users, VIRTE tried to minimize program license costs with the vision that of being able to hand out CD applications to anyone that wanted them. These requirements led to the selection of the government owned JointSAF as the simulation environment. OpenFlight was selected as the visualization database standard and allowed each of the development teams to choose their own tools. VIRTE's goals in this domain came closest to the vision with the VELCAC which uses a commercial gaming engine, Gamebryo (formerly NetImmerse). VIRTE Demonstrations II and III will all use Gamebryo.

Early prototypes were crucial to the program. In a two year span, VIRTE went from Demonstration I concepts to deployable prototypes. VIRTE developers employed a Spiral Development process and held a series of Intermediate Feasibility Experiments (IFEs). The four IFEs were events where developers deployed their latest configurations in realistic testing environments with potential users. All of the VIRTE simulations are PC based and interoperable in the same virtual battlespace. They have some unique differences based on customer requirements. The VEAAAV, which is a school house training prototype, has extensive replication of the physical layout of the crew stations. The VELCAC uses almost all virtual displays, with the exception of the "throttle" which has a unique feel and gives important haptic feedback to the craftmaster. The VEHelo is the only VIRTE Demo I simulator that uses a HMD in the deployed configuration. The VEHelo uses a Head Mounted Display (HMD) which combines the virtual environment and live video. Each of these approaches has tradeoffs and these are brought out in the Training Effectiveness Evaluation.

VIRTE Demonstration II, Close Quarters Battle for Military Operations in Urban Terrain (CQB for MOUT) has a much greater focus on "traditional" Virtual Environments. The most unique aspect of Demo II is that it puts a soldier directly into the virtual environment using his entire body and weapon as a means of interaction. Unlike First Person Shooter (FPS) games, which use keyboard commands or a joystick, VIRTE is prototyping technology which allows the user to naturally interact with the virtual environment either walking in a tracked space or walking in place. This level of interaction with the Virtual Environment presents exceptional challenges for Human Factors. For example, what is the best way to let a person know that they have put their hand through a virtual wall? How do we immerse an entire fire team and allow them to have natural interactions? In Demo II, the CQB training system, VIRTE is focusing significant investment into real time full body tracking, collision detection, distributed audio environments, and distributed Virtual Environments. The remainder of this paper discusses the process the VIRTE team has implemented to ensure that the training provided by each VE system provides a measurable level of performance enhancement on the real world tasks they are designed to support.

Training Effectiveness Evaluation (TEE)

In order to better understand the process implemented by the VIRTE team for ensuring training enhancement, the individual components comprising this user-centric design process were categorized as elements of a broader, Training Effectiveness Evaluation (TEE) effort. The philosophy underlying the development of a TEE plan includes:

- Up Front Analyses
    - Provide foundation for developing systems
- Iterative Integrated Feasibility Experiments
    - Provide iterative feedback for system developers
- Back end Analyses

– Final evaluation
– Lessons Learned

These three general elements can be operationalized as six component ‘building blocks’ which are:

- Task Analysis
- Human Computer Interface evaluation
- Team Performance
- System Usability
- VE User Considerations
- Training Transfer

The manner in which these pieces combine is somewhat fluid. In a broad sense, these pieces can be grouped into three general phases. Phase one consists of the Task Analysis and the Human Computer Interaction evaluation. The first element, the *Task Analysis (TA)*, seeks to identify the training objectives and the scenario elements that must be included in a simulation to support these objectives as well as providing an assessment of whether currently available training can be modified so that a technology solution is not necessary. For TEE purposes, a contextual task analysis is performed, which focuses on the behavioral aspects of a task as performed in a given operational setting, resulting in an understanding of the general structure and flow of task activities (Mayhew, 1999; Nielsen, 1993; Wixon & Wilson, 1997). The second element *Human-Computer Interactions (HCI)* focuses on identifying the requirements for sensory modality integration, as well as evaluating current hardware/software technologies supporting these interactions, and providing guidance for integrating them into a simulation. Once these efforts are completed, the information is then passed on to the Simulation Developers, who may then commence their efforts.

The second phase involves an iterative evaluation process, in which the system is evaluated at key points within the development lifecycle. These points, termed Integrated Feasibility Experiments are points at which the TEE team can evaluate progress along the dimensions determined in phase 1. There are two key portions to this phase. The first is *System Usability*, which evaluates how accommodating the overall VE design is for use by the layperson (Stanney, Mollaghasemi, & Reeves, 2000) and provides redesign recommendations to the developers. The second is *VE User Considerations,* which addresses both evaluations of the side effects encountered during exposure to VE, as well as aftereffects arising following this exposure.

The final phase focuses on performing a system-wide evaluation, and then documenting these findings in a ‘lessons learned’ format. Since all the training systems under development for VIRTE focus on crew or team-based application, the first component of this effort is Team *Performance*, which addresses how well a simulation supports team-based activities (Cannon-Bowers, & Salas, 1998). Perhaps the most important element of TEE, which validates all the efforts described thus far, is an evaluation of the degree to which *Training Transfers* from the VE to the real world scenario (Carretta & Dunlap, 1998; Lathan, Tracey, Sebrechts, Clawson, & Higgins, 2002; Waller, Hunt, & Knapp, 1998). Figure 2 provides an overview of how these components can be integrated into a comprehensive TEE plan.

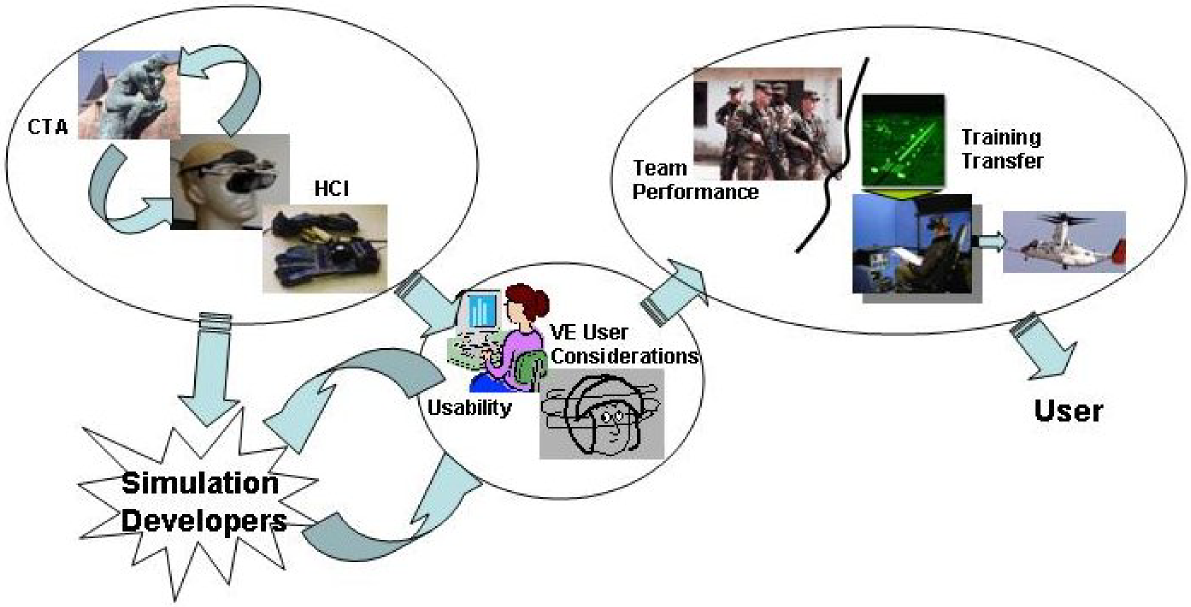


Figure 2: Schematic of elements comprising the VIRTE Training Effectiveness Evaluation, and the relationship/flow structure linking them.

As Figure 2 suggests, TEE is not simply a process that is performed once, at any single stage in the design of a VE simulation. Rather, as with traditional usability engineering practices (Nielsen, 1993), it must be interwoven throughout the entire design process. Moreover, each of the proposed TEE elements is intertwined with the others. Ultimately, the six aspects of TEE should combine to form a seamless, continuous, evaluation of the overall VE training simulation, leading to a well thought out, well-designed system that will enhance training while reducing the cost of providing this training. The research objectives of VIRTE were structured so that, while each VE platform utilized aspects of each TEE component, each platform was also singled out to highlight, from the science and technology perspective, issues and applications within these elements.

Task Analysis

One of the most basic challenges that any VE developer faces, prior to designing a new VE simulation is *How to design a system that will support the users' training needs?* The commonly accepted approach for answering this question is to conduct a task analysis. While there are many types of task analysis (Chapman, Schrage, & Smith, 2001), it is possible to select one basic methodology by understanding the characteristics of the user community. VIRTE is concerned with providing experts with training on abstract concepts rather than with training novices to perform specific perceptuomotor tasks, hence a *cognitive task analysis* (CTA).

Through this effort, training objectives and evaluation metrics were developed for each platform. Considering the training of the LCC community, it was determined early on that the VELCAC system's primary use would be as a teaching supplement to acquaint highly skilled crews with a new cockpit instrumentation layout and corresponding upgraded features (the Service Life Extension Program, or SLEP, LCAC configuration). The resultant CTA identified those features of VELCAC that would best complement the SLEP differences course, a two-week training course designed to acquaint crewmembers with the new LCAC features. The key training objective for the VE AAAV was identified as being the provision of a gunnery training system. Hence, the task analysis focused on techniques for providing real time and after action feedback to enhance speed and accuracy of this task. The focus of the VE Helo system was on providing enhanced training for visualizing spatial relationships in preparation for low level, Nap of Earth (NOE) flight. Consequently, the VE Helo task analysis focused on identifying the cognitive components of route learning and navigation. Finally, the intended use of the CQB training system is to provide a means for maintaining skill proficiency during extended deployments, as well to provide an interactive, three-dimensional mission rehearsal system. The CQB task analysis focused not only on core skills and cognitive components of basic CQB maneuvers, but also on understanding tactics and strategies that the Opposing Forces (OpFor) might use.

Human Computer Interaction

The overall goal of the HCI evaluation is to identify, based on the CTA, which modalities (haptics, visual, sound, etc.) must be represented within the VE to in order to provide an effective training experience. Since the bulk of the VIRTE systems are vehicle- based, the two primary senses under consideration are vision and haptics. Although much work has been done exploring the visual modality, the integration of haptics within a VE simulation represents a unique challenge because, unlike vision (and audition), the haptic system provides both sensation/perception and the means to manipulate.

The bulk of VIRTE's HCI effort focused on the VELCAC system. Several sources of information were examined during this evaluation. These included a requirements document, user profiles, results from user interviews, the TA, field observations, and knowledge of the operational environment. In addition, a literature review examining performance issues relating to visual displays and haptic systems was conducted. The analysis of this information provided a set of criteria by which decisions regarding visual and haptic displays could be made. Critical information derived from these sources suggested that of the three crew positions under consideration, only the Craftmaster (the crewmember who flies the LCAC) would need physical controls (yoke, foot pedals, and throttle) in the VELCAC and a wide field of view, in order to maintain adequate situational awareness while piloting the (virtual) craft. The other two crewmembers (the Navigator, who monitors instant position, and the Engineer, who monitors LCAC system function) spend the majority of their time observing and interacting with touch screen interfaces on the instrument console. Consequently, the final version of VELCAC utilized a panoramic-like, flat-screen based visual display. The Craftmaster position was provided with the actual controls interfaces, while the other two positions were provided with an interface to their instruments based on using a mouse to point and select specific options.

Usability

System usability focuses on evaluating how accommodating, intuitive and easy-to-use software and hardware are for use by the layperson. The assessment of a system's usability typically involves an initial expert evaluation of the system to identify any design weaknesses in the prototype system, which includes an observational assessment of users interacting with the VE, an analysis of these data to identify which features are less-than optimal and why and, the development of a set of ranked and ordered redesign recommendations for system developers to follow. As should be apparent Usability evaluation is most useful when it is incorporated as part and parcel of an iterative design and development process.

As an example of how this iterative process works, during each IFE the usability assessment of the VELCAC began with an expert evaluation of the current version of VELCAC, conducted by two usability engineers. This assessment was then followed by user observation, by these same experts, of LCAC crewmembers interacting with various SLEP VELCAC stations. The results of these efforts were used to provide redesign recommendations, as well as to establish/validate usability metrics and their associated acceptability criteria for future usability testing of SLEP VELCAC as more mature iterations of the system are developed.

VE User Considerations

Typically, the impact of Virtual Environments on users has been typified in terms of a collective group of symptoms known as Cybersickness. Operationally, the majority of the research into these symptoms has focused on acute symptoms, such as eyestrain and blurred vision, rather than on a more holistic approach, which would focus on impact on warfighter readiness. Yet, as the military continues its push towards using VE systems to provide just in time training (Cohn, Muth, Schmorrow, Brendley & Hillson, 2002), it is becoming increasingly evident that this level of research will be the one most critical for evaluating the utility of using VE systems. Currently, the military has restrictions delineating the amount of time an individual must wait between exposure to simulated training and participating in the actual real world effort.

For the purposes of VIRTE, whose simulators target Naval forces and may ultimately be placed aboard ship (Cohn et al, 2002), a critical concern is the sensory decoupling that arises between the visually indicated motion reference frame, provided through immersion within the VE, and the physically indicated motion reference frame, provided through the (moving) platform of the ship. Such sensory discordance may be particularly troublesome, both during training as well as following training. Preliminary results (Cohn et al., 2002) suggest that in this situation side effects, such as motion sickness will result. VIRTE is actively pursing two solutions to this sensory discordance challenge. The first focuses on developing technology to re-introduce the coupling between physically sensed and visually indicated motion. The physically indicated motion of the ship is locked to the motion of the entire framework of the virtual scene being displayed through the VE in such a manner that a trainee experiences both the intended, virtual motion, as well as a degree of motion recoupling. Preliminary results suggest that this approach holds promise (Brendley, Marti, Cohn & DiZio, 2002). At the same time, it is crucial to quantify the impact these decouplings have on warfighter readiness. Consequently, VIRTE is developing a Warfighter Readiness Toolkit that combines cognitive, perceptual and physiological evaluation tools that can detect subtle changes that might lead to deleterious side effects that could impact warfighter readiness.

Team Performance and Training Transfer

Ultimately, each of the previously mentioned efforts are useful only to the extent that they are able to support the primary goal of developing *any* VE simulation, namely, to provide a level of training that translates to enhanced performance of the real world task being simulated (Lathan et al., 2002). Despite the fact that VE systems continue to be pushed into the training application domain, there are few convincing illustrations of the transfer of complex skills from VE training to their corresponding real-world applications (Cohn, Helmick, Meyers & Burns, 2000). A critical limiting factor in most such efforts is that they fail to utilize adequate performance metrics, relying instead on users' anecdotal evaluations or perceptions of training utility rather than statistically determined measures. Thus most results are difficult to interpret both in terms of the underlying *processes* that lead to either positive or negative transfer, as well as in terms of the meaning of *outcome*.

The most common approach for assessing training transfer is to compare performance between two groups of trainees, an experimental group that receives simulator training and a control group that receives all of its training in the real world, along some predefined set of metrics (Boldovici, 1987; Cohn et al., 2000; Murdock, 1957). It is important to note that while it may be possible, in a general sense, to develop standard approaches for this effort, ultimately, the metrics thus developed will be context specific. Thus, for each type of VE system, it is critical to identify differences in process and outcome behaviors between individuals who receive VE training and those who do not receive that training when performing operational tasks in the real world environment.

Importantly, for purposes of a transfer study, each of the real world platforms being simulated through VIRTE involves interactions between team members. Consequently, measures of changes in team processes (e.g., exchange of information, backup behaviors during times of high workload) serve as a critical measure of training effective and directly impact the degree to which training will transfer. At the same time, more task specific, outcome measures should also improve, such as improved timing (e.g., at checkpoints and way-points) and accuracy (e.g., less discrepancy between actual and planned speeds and course), which should occur if there is positive transfer of training from the VE to the real world.

Summary

While many elements of Virtual Environment development are platform specific, it is possible to distill a set of principles, which if followed, should translate to a more effective training system. These principles can be implemented early on, as part of the task analysis and human computer interaction evaluation efforts and can continue, throughout the development lifecycle as a usability and user consideration assessment. Ultimately, the success of these efforts will be established through a comprehensive training transfer study.

References

Boldovici, J. A. (1987). Measuring transfer in military settings. In S. M. Cormier (Ed.), Transfer of learning: contemporary research and learning (pp.239-260). San Diego: Academic Press, Inc.

Brendley, K., Maarti, J., Cohn, J.V., & DiZio, P. (2002). Motion Coupling in Virtual Environments (MOCOVE). The 46th Annual Human Factors and Ergonomics Society Meeting. Baltimore, MD, September 29-October 4, 2002.

Cannon-Bowers, J.A., & Salas, E. (1998). Team performance and training in complex environments: Recent findings from applied research. Current Directions in Psychological Science, 7: 83-87.

Carretta, T. R., & Dunlap, R.D. (1998). Transfer of Effectiveness in Flight Simulation: 1986 to 1997, U.S. Air Force Research Laboratory, NTIS.

Chipman, S.F., Schraagen, J.M., & Shalin, V.L. (2000). Introduction to cognitive task analysis. In J. Maarten, Schraagen, S. F. Chipman, & V.L. Shalin (Eds.), Cognitive task analysis (pp. 3-23). Mahwah, NJ: Lawrence Erlbaum Associates.

Cohn, J.V., Helmick, J., Meyers, C., & Burns, J. (2000). Training-transfer guidelines for virtual environments (VE). Presented at 22nd Annual Interservice/Industry Training, Simulation and Education Conference, Orlando FL.

Cohn, J.V., Breaux, R.B., Nguyen, L.K., & Schmorrow, D. (2001). Designing and Implementing an Effective Virtual Reality Trainer. Tutorial presented at IEEE VR2001 Conference. Pacifico Yokohama Conference Center, Yokohama, Japan.

Cohn, J. V., Muth, E., Schmorrow, D, Brendley, K., & Hillson, R. (2002). Reducing Negative Effects from Virtual Environments: Implications for Just-In-Time Training Presented at NATO Human Factors and Medicine Symposium on Spatial Disorientation in Military Vehicles: Causes, Consequences and Cures, La Caruna, Spain.

Department of the Navy, Headquarters United States Marine Corps, *Expeditionary Maneuver Warfare Marine Corps Capstone Concept*, 10 November 2001

Lathan, C.E., Tracey, M.R., Sebrechts, M.M., Clawson, D.M., & Higgins, G.A. (2002). Using virtual environments as training simulators: Measuring transfer. In K.M. Stanney (Ed.), Handbook of virtual environments: Design, implementation, and applications (pp. 403-414). Mahwah: NJ: Lawrence Erlbaum Associates.

Mayhew, D. J. (1999). The usability engineering lifecycle: A practitioner's handbook for user interface design. San Francisco, CA: Morgan Kaufmann.

Murdock, B.B. (1957). Transfer design and formulas. Psychological Bulletin, 54(4): 312-326.

Nielsen, J. (1993). Usability engineering. Boston: Academic Press.

Stanney, K.M., Mollaghasemi, M., & Reeves, L. (2000). Development of MAUVE, the Multi-Criteria Assessment of Usability for Virtual Environments System. Final Report, Contract No. N61339-99-C-0098, Orlando, FL: Naval Air Warfare Center – Training Systems Division, 8/00.

Waller, D., Hunt, E., & Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. Presence, 7: 129-143.

Wixon, D., & Wilson, C. (1997). The usability engineering framework for product design and evaluation. In M. Helander (Ed.), Handbook of human-computer interaction (pp. 653–688). Amsterdam, Netherlands: North-Holland.

Training in Peacekeeping Operations Using Virtual Environments

Dr. R. Bowen Loftin, Dr. Mark W. Scerbo, Dr. Rick McKenzie, Ms. Jean M. Catanzaro, Mr. Nathan R. Bailey, Mr. Mark A. Phillips, and Dr. Gaye Perry[1]
Virginia Modeling, Analysis, and Simulation Center
Old Dominion University
Suffolk, VA 23435 USA

Summary

The present paper describes two studies aimed at evaluating Virtual Environment (VE) technology for training individuals to perform military checkpoint duty. Participants stood guard at a fictitious base in which simulated drivers in vehicles approached seeking entrance. Participants inspected each vehicle, interacted with the drivers, verified their identification, and made a decision to allow the driver to enter the base, detain the vehicle, or asked the driver to turn around and leave. The first experiment was conducted in a CAVE environment with stereoscopic visual and auditory displays, participant tracking, and voice recognition. The second experiment provided the same training on a desktop system. The results of both studies showed that participants learned quite effectively with either interface, but that overall levels of performance were better with the fully immersive VE. These findings suggest that VE technology holds promise for activities that are more like experience-based training and which place a greater emphasis on social interaction skills.

Introduction

Military training has been traditionally aimed at preparing soldiers to apply doctrine and to react instinctively to accomplish objectives. This type of training is necessary for soldiers to win wars and to minimize casualties and collateral damage. Unfortunately, these are the same soldiers who become ambassadors, peacekeepers, and police in a disrupted state where the complexities of the environment are so great that instinctive or skill-based behavior is simply not enough to cope with unexpected and complex situations. Too often, soldiers are faced with difficult and politically-sensitive decisions for which they have received no training at all.

During the past 20 years, the United States has engaged in two wars but has been a participant in nearly thirty major peacekeeping operations. Further, since most military operations today are intensely scrutinized by the news media, it has become evident that the actions of even the most junior members of a military unit may profoundly impact world opinion and affect the most senior levels of leadership.

This was made abundantly clear in the recent war in Iraq. On March 29, 2003, a suicide bomber at a checkpoint near Najaf driving a taxicab feigned engine trouble. When soldiers approached to inspect the situation, the driver blew up the vehicle. Four soldiers were killed. Within 48 hours, another vehicle near Najaf failed to heed warnings to stop as it approached a checkpoint. After several unsuccessful attempts to try and slow the vehicle, the soldiers fired into the vehicle, killing seven women and children. Each of these incidents made headlines and required military commanders to scrutinize their rules of engagement. Thus, proper training of military personnel, at all levels has never been more crucial.

A recent review of military applications of virtual environments (VEs) indicates that most efforts have focused on training teams, leadership skills, mission rehearsal, and navigational skills (Knerr, Breaux, Goldberg, & Thurman, 2002; Pew & Mavor, 1998). Although the potential of VE technology for addressing the interpersonal skills needed in many military activities has certainly been acknowledged, few systems have actually been developed to meet this need. Accordingly, the present paper describes the application of VE technology to a novel task, the military checkpoint.

---

[1] University of Houston

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

The primary objective was to recreate the kind of experiences a military guard would encounter while standing watch and to determine whether a VE would prove to be an effective learning medium. Toward this end, computer-controlled virtual humans and live human participants took part in a peacekeeping task whereby various skill-based activities were trained and evaluated in a VE. The task used was a checkpoint operation in a typical third world urban area. Simulated drivers in vehicles would approach a checkpoint seeking entrance to a fictitious military base. Trainees would inspect each vehicle, interact with the drivers, and ask for proper identification. The driver would produce an ID card and the trainee had to verify that it was valid. The scenario would end when the trainee made a decision to allow the driver to enter the base, to pull over the vehicle, or ask the driver to turn around and leave.

The training addressed six specific objectives (see below). In the first study, the training took place in an immersive environment using CAVE (CAVE Automatic Virtual Environment) technology. The system incorporated speech recognition software with a focused natural language interface. Trainee movements within the environment were monitored by an Ascension Flock of Birds magnetic tracking system. This tracking information was provided to the virtual human agents in the environment. The technology permitted a high level of interaction between trainees and the human models. These virtual agents answered questions, knew where the trainees were in the environment, and replied while looking the trainees "in the eye."

In a typical scenario, a trainee would approach the car and ask the virtual driver for identification. The trainee's virtual partner provided cover for the trainee during the identity check. The driver produced an ID card and the trainee verified that it was appropriate.

Two groups of individuals participated. The first group participated in one session, received feedback on their performance, and then returned for a second session. It was expected that these individuals would perform better (i.e., make fewer errors) in the second session if the VE was an effective training medium. A second group of individuals participated in only a single session. These participants were trained with the same scenarios that the first group performed in their second session. This second group served as a control for the specific scenarios performed by the first group on their second session. Thus, if the participants in the first session truly benefited from their training, one would expect their performance on the second session to be superior to that of the second group who only performed a single session. Further, performance levels for Group 2 should be similar to those of Group 1 on their initial session.

The second study was a replication of the first experiment with one important difference. The scenarios were presented on a desktop VE system. A new interface was created to allow the trainees to navigate and inspect the vehicles. This study was intended to provide a comparison between the immersive CAVE and desktop VE.

The Training System

Virtual Human Agents

Virtual human agents were created with Jack Tool Kit, a 3D modeling environment with support for high degrees of freedom human models. These models are typically used to evaluate ergonomic factors pertaining to the modeled environment. The human models within Jack were selected for this project because of the range of dynamic motion available. Jack includes utilities for locomotion, head and eye movement, arm and leg movements, and movement of all joints. The extent of motion of the human models is always within the physical constraints of selectable human body types. As a result, one is assured of gestures and positions that are within the realm of possibility, given the particular human in a particular environment.

Behaviors in Jack are supported through layers of interfaces with decreasing complexity. At the lowest level, rotations and translations of 68 joints in the human figure are supported. Above this layer are a number of primitives that control movement of individual human body parts such as Move Arm, Move Head, Bend Torso, Rotate Pelvis, etc. These primitives are combined to create an executable behavior in the Jack agent instantiated for the target application. A network of these executable behaviors provides the activities and reactions that the agent will exhibit during part or possibly throughout an entire scenario. The network consists of basic transition nodes as well as nodes that can execute in parallel. Thus, the behavioral network is called a Parallel Transition Network (PatNet; Badler et al., 2000). Another layer of generalized capability called Parameterized Action Representation (PAR) is available in Jack that supports both natural language commands and automatic behavioral animation (Badler et al., 2000).

Training Scenarios

The present set of experiments examined the application of VE technology to a military checkpoint task. The primary objective was to reproduce the kind of experiences a military guard would encounter while standing watch. Thus, the task, setting, and virtual characters were created to match typical checkpoint conditions as closely as possible.

The checkpoint locale was recreated from the U.S. Marine training town in Quantico, VA. A pre-existing Quantico model was updated by remodeling existing structures for improved real-time performance and by applying texture map created from photographs obtained at the site. Scene graph construction and rendering was done with VrTool. There are two scene graphs, one in VrTool and the other in Jack. VrTool's scene graph is what is actually seen and rendered on the screen, whereas Jack's is used internally for dynamic character animation calculations.

The task required the creation of many distinct training scenarios. In actuality, the process of manning a checkpoint can be a highly repetitive, mundane activity. Rarely does anything out of the ordinary occur. Accordingly, the training scenarios were designed from this perspective. Specifically, a general or neutral scenario was created which begins when a vehicle approaches and stops at the checkpoint. The trainee inspects the vehicle and asks the driver, a virtual human intelligent agent, for identification. The driver produces an ID card. The trainee verifies that it is valid and, if so, allows the driver to pass. The trainee's partner (another virtual agent) provides cover for the trainee during the interaction.

In order for the training experience to more closely reproduce the true conditions of this activity, each neutral interaction had to be unique. Thus, a pool of neutral scenarios were generated that varied in vehicle type, vehicle color, driver's sex, skin color, hair color, and shirt color. Although the characteristics of the neutral scenarios varied, they all unfolded in the same manner. In this regard, every attempt was made to create a training environment that would reproduce the *experience* of standing watch at a checkpoint.

The participants' ability to follow protocol and exercise judgment was examined by including a variety of critical scenarios. The critical scenarios appeared at random intervals throughout the training session and unfolded without any cues to distinguish them from the neutral scenarios. Specifically, these scenarios addressed the following training objectives:

a) the ability to handle matters of situational urgency according to procedure;
b) the ability to resist social pressures that conflict with procedure;
c) the ability to recall and identify vehicles, people, and license plates from a predefined target list;
d) the ability to perceive inappropriate objects/contraband or the absence of required information;
e) the ability to maintain situation awareness.

Thus, for example, in one scenario an ambulance arrives without proper authorization via radio alert and the driver advises the trainee that he does not have time to go through the normal identification verification routine because he has an injured passenger. The trainee is responsible for the security of the base and must follow proper procedure and perform an identification check on both the driver and passenger even if confronted with an urgent situation. In another type of critical scenario, the trainee was presented with specific information about a vehicle that he/she needed to remember and was instructed to watch out for during his/her shift. Other scenarios required the trainees to detect a missing base sticker or identify the presence of contraband items (see Figure 1). The trainees were also evaluated on their ability to maintain situational awareness by attending to two events simultaneously. Specifically, in one scenario the driver of a second vehicle awaiting entry to the base exhibits suspicious behavior. The trainee must interact with the driver in the first vehicle while monitoring the activity in a second vehicle. In this particular scenario, if the participant failed to ask for backup from their virtual partner, the driver in the second vehicle would pull out a gun and fire. Although the scenarios are mainly skill-based, they are still representative of the kinds of judgment a checkpoint guard must make.

Experiment 1: Fully Immersive CAVE Environment

Virtual Environment Implementation

The VE interface used in the first study was the CAVE (CAVE Automatic Virtual Environment).

Figure 1. Example of a vehicle approaching the checkpoint. Note the missing license plate.

The system configuration for the first experiment is illustrated in Figure 2. There are three main computing systems connected through a 100-mbps network switch as described below:

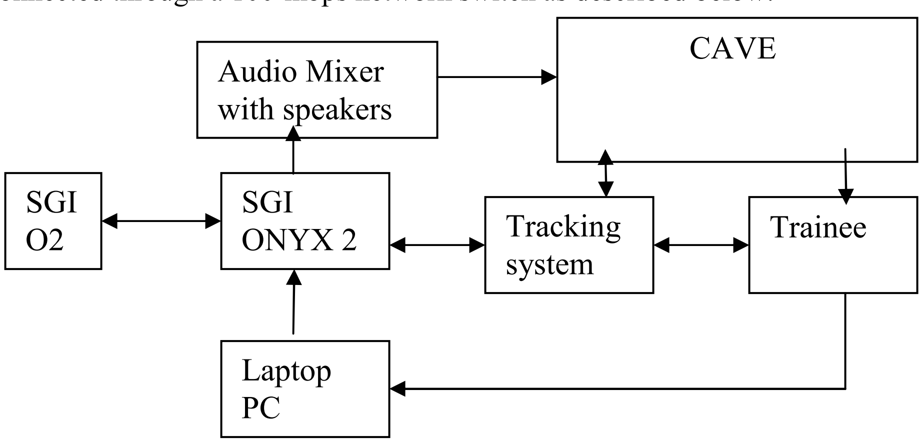


Figure 2. Experiment 1 hardware configuration.

- An SGI ONYX 2 computer was used to display the application in the CAVE, provide the sound playback, and read the information from the tracking devices. This computer used VrTool, TrackD, Jack, Python, Open Inventor, and IRIX 6.5.
- An SGI O2 computer was used as the experiment main console. From that machine one could launch the application and have override controls during the simulation. This computer used IRIX 6.5, Motif, and buttonfly.
- A PC computer was also used for the voice recognition software and to communicate the information to the SGI ONYX 2 through a network socket. This computer used Windows 2000, IBM ViaVoice, and VrSpeech.

Images were presented on two 10x10 ft walls of the CAVE with a resolution of 1280x1024. The images were viewed stereoscopically with LCD CrystalEyes stereo shutter glasses. Positional tracking was provided through Ascension Technology's Flock of Birds software, a six-degree-of-freedom (6DOF) tracker able to track one to four sensors simultaneously. A single head sensor was attached to the CrystalEyes LCD shutter glasses. The participant wore a wireless headset microphone to communicate the voice commands to the PC running IBM's Via Voice Recognition software.

Audio Elements

The IBM ViaVoice Speech Recognition required two components—a grammar and a dictionary. The most difficult challenge concerned creating the grammar. The SRCL (Speech Recognition Control Language) used was a particular type of the BNF (Backus-Naur Form) generic grammar representation (IBM Corporation 1997). It supports substitutions and repetition and can generate very complicated sentences while at the same time addressing a wide selection of the possible commands. The dictionary provides the software with pronunciations for each word to be recognized. In this study, it was designed for an East Coast USA accent. For each word, it was necessary to say the word to record it. The software would then convert it into a corresponding 'baseform' representation according to its lexeme (spelling).

Communication from the laptop program to the main program running on the SGI was established using sockets. It connected to VrSpeech (a component of VrTool) that is designed specifically to receive ASCII strings via socket communication.

Audio files were created for the virtual humans and for sound effects. For the speech files, male and female voices were recorded for a variety of phrases such as: "Here's my ID, I don't have a pass, I don't understand," etc. For the neutral scenarios, a group of four male voices and four female voices were recorded, using an identical set of scripts for each individual. Because only eight unique voices were recorded for the neutral scenarios, there was some repetition of the individual voices. For each of the critical scenarios, a unique script was written and recorded. For the critical scenarios, 12 male and seven female voices were used. Again, in some cases, the same voice was used in more than one critical scenario. However, when repetition did occur, it took place in critical scenarios that were presented in different experimental sessions.

Each voice was recorded in mono at a 22.1 kHz sampling rate. Background and other supplemental audio sounds including gunfire, airplane flybys, wind, and the approach of a car on a gravel surface, were created using a combination of existing sound samples and environmental sound recordings. A noise reduction algorithm was used to eliminate unwanted noise (including hiss, clicks and pops), for each of the audio files. The files were converted to Audio Interchange File Format Version C (.AIFFC) for final presentation in the CAVE environment

.All sounds were displayed in the CAVE environment through a standard four-channel soundboard, though only two of those channels were used for the auditory presentations in this experiment. With the experimental participant facing forward in the CAVE, the left and right speakers were placed at approximately 225 and 315 degrees, respectively. The speakers were mounted on speaker stands at an elevation of approximately five feet. Throughout the majority of the experimental session, the speech files and background samples did not exceed 85dB.

Experimental Method

Participants. Thirty-two undergraduate students from Old Dominion University with normal or corrected-to-normal vision participated in the study. They were offered either (a) four hours of extra credit or (b) $30 as compensation for their time. This population was chosen because they were representative of the type of individual who would likely be assigned to guard duty in the military. Participants who were predisposed to simulator sickness as indicated by the Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993) were excluded from participating. In addition, individuals who reported previous experience with checkpoint duty were excluded from the analyses and four were replaced due to system malfunctions.

General procedure. Participants were all run individually. They were asked to complete a background survey addressing demographics and military experience and then review a three-page training manual outlining procedures for their shift as a checkpoint guard. They then watched a 7.5-minute video that provided information about their checkpoint location, existing threat conditions, proper radio call procedures, and the need to maintain vigilance and alertness. Participants were told their job was to assess the vehicle and all of its occupants and render a decision as to whether they could enter the base. They were also told that they were part of a team and that they had a virtual teammate who would provide cover during their interactions with the drivers.

They were then taken to the CAVE and fitted with their equipment: an inert pistol and holster as well as a walkie-talkie to communicate with the base. Next, the participants were given a log sheet depicting a time log of events that occurred on "the previous shift" and "Be On the Lookout" (BOL) information. The BOLs on the log sheet described events the participant was required to remember and look for throughout the session. Additional BOL events were presented aurally during the checkpoint task. Next, the participants were given a chance to familiarize themselves with a typical scenario and the equipment used to interact with the virtual people in the scenario. They were given ample time to repeat this process until they were comfortable with the task.

A video camera was used to record each participant's performance during the scenarios. The camera was positioned to record a fixed image of the participant within the CAVE. A pair of LCD CrystalEyes shutter glasses was taped over the camera lens to record a single image from the stereoscopic display.

Experimental design. Training was assessed in two ways. Participants were randomly assigned to one of two groups. Those in the first group performed a 45-minute shift, received feedback on that shift, and then performed a second 45-minute shift. Performance was compared between the two shifts. Participants in the second group performed only a single session. The performance of these participants was compared to the performance obtained from the second session of the participants in Group 1.

Procedure: Group 1. The first experimental session contained 23 neutral scenarios and 12 critical scenarios, and each interaction took 1 to 2 minutes, on average, to execute. Upon completion of the first session, participants were given an after-action review (AAR) in which they received corrective feedback regarding the nature of their errors and the proper resolution of those errors. The videotape of the participant's performance was replayed if necessary. The participants were also given a second questionnaire to assess postexperimental levels of stress and symptoms of simulator sickness. These participants returned 48 hours later and performed a second 45-minute shift. The second experimental session contained 22 neutral scenarios and a different set of 11 critical scenarios that were conceptually similar to those from the first session and addressed the same training objectives.

Procedure: Group 2. Participants in Group 2 performed only one session. Specifically, they performed the same session that Group 1 performed on their second shift. Group 2 was included because it could be argued that the unique scenarios contained in session 2 were easier than those in session 1. If that were true, one would expect the performance of Group 2 to be similar to that of Group1 on session 2. On the other hand, if the participants in Group 1 truly benefited from their training, one would expect their performance on their second session to be superior to that of Group 2 who only performed a single session. Further, the level of performance of Group 2 should be similar to that of Group 1 on their *initial* session. Other than the set of scenarios, the experimental procedures for Group 2 replicated those used in the first session with Group 1.

Results

Performance was assessed by the total number of errors made by each participant on each scenario type. An alpha level of .05 was used for all statistical comparisons. Tukey post hoc tests were used to analyze differences among the means.

Group 1. Performance data were analyzed in two ways. The first analysis addressed performance differences between session 1 and session 2 for only the Group 1 participants. The number of errors for critical and neutral scenarios in both sessions was compared using a 2 x 2 within-subjects ANOVA. A significant effect for scenario type indicated that participants committed significantly more errors on critical scenarios ($M = 4.21$, $SD = 5.17$) than on neutral scenarios ($M = .80$, $SD = 1.81$), $F\ (1, 15) = 34.31$, $p < .01$. In addition, a significant session effect indicated that more errors were made in the first session ($M = 2.94$, $SD = 4.14$) than in the second session ($M = .92$, $SD = 2.67$). $F\ (1, 15) = 12.93$, $p < .05$. Further, there was a significant interaction between session and scenario type, $F(1,15) = 12.14$, $p < .05$. The nature of that interaction is shown in Table 1. As can be seen in the table, the mean number of errors dropped considerably from session 1 to session 2 for both types of scenarios, but the decline was slightly more pronounced for the critical scenarios.

Table 1.
Mean Total Errors for Critical and Neutral Scenarios in Each Session.

	Session1	Session 2
Critical Scenarios	6.36 (6.0)	1.87 (4.51)
Neutral Scenarios	1.15 (2.53)	.44 (1.1)

Standard deviations appear in parentheses.

Group 2. The second analysis compared performance between the participants in Group 2 and Group 1 in their second session. The analysis followed the same format as that described above. Thus, in the first analysis the mean number of errors for critical and neutral scenarios was examined for participants in Groups 1 and 2. This analysis used a 2 x 2 mixed-factor ANOVA with group analyzed as a between-subjects factor and scenario type analyzed as a within-subjects factor.

As hypothesized, Group 1 participants made significantly fewer errors (M = .92, SD = 2.67) in their second session than participants in Group 2 (M = 2.31, SD = 2.62), $F\ (1, 30) = 9.58$, $p < .01$. A main effect for scenario type was also found, $F\ (1,30) = 52.44$, $p < .001$. Participants committed more errors in critical scenarios (M = 3.52, SD = .3.96) than neutral scenarios (M = .66, SD = 1.54). Again, there was a significant interaction between condition and scenario type, $F\ (1, 30) = 11.16$, $p < .005$. The interaction is shown in

Table 2. As can be seen in the table, the participants in Group 2 made more errors on both types of scenarios, but their poorest performance was on the critical scenarios.

Table 2.
Mean Number of Errors for Critical and Neutral Scenarios in Each Group.

	Group 1	Group 2
Critical Scenarios	1.87 (4.51)	5.17 (3.41)
Neutral Scenarios	.44 (1.1)	.88 (1.98)

Standard deviations appear in parentheses.

The better performance of participants in Group 1 during session 2 over those of Group 2 indicates that they benefited from their training experience. If the scenarios used in the second session were "easier" than those used in the first session, one would have expected both groups to perform at the same level. That did not happen. This argument is also supported by comparing the performance of participants in Group 1, session 1 with those of Group 2. An independent t-test used as a manipulation check indicated there was no significant difference in errors between Group 1, session 1 and Group 2, $t(30) = -.34$, $p > .05$. Thus, the scenarios used in each session can be considered equivalent.

Experiment 2: Desktop Interactive Training

Experiment 2 was intended to provide a comparison between the immersive CAVE and VE desktop interfaces. In an effort to examine the benefits of an immersive environment for this task, all other parameters were held constant to the greatest extent possible. As a result, there were few differences in the software, scenarios, and protocols. Specific differences between the two interfaces are described below.

Desktop Virtual Environment Implementation

There were two main differences in the hardware configurations between Experiments 1 and 2. First, the SGI ONYX 2 Image Generator (IG) was replaced with an SGI Octane desktop computer. Also, the images were not displayed stereoscopically and instead were presented on an 18-inch Sony flat panel display. Therefore, no shutter glasses were used. Second, there was no positional tracking with the desktop system. Instead, a new interface had to be developed to allow the participant to navigate within the scene and inspect the vehicles (see Khan, 2002).

The Graphical User Interface (GUI) was placed over the rendered scenarios as shown in the Figure 2. The GUI had two drop-down menu options in the toolbar: menu (play, pause, speech toggle, or quit) and zoom (zoom in, zoom out, or zoom reset). There were also push buttons that provided features such as *walk to driver*, *walk to neutral*, *look at driver*, and *change view*. The first two allowed the avatar to walk to a location that is an offset from the driver and to walk back to a pre-determined neutral location. *Look at driver* bends the avatar's torso so that the driver comes into his line of sight. Change view simply toggled between the original camera view (overlooking the checkpoint scene) and the avatar's view. In addition, the GUI had buttons (large directional triangles) to pan the camera left, right, up, and down, and to reset it to a default position. Using these buttons, a trainee could inspect the front and back seats of a vehicle.

A second GUI (see Figure 3) was required for the experimenter's computer to execute driver and passenger responses and to force commands that the voice recognition software failed to recognize. This coordination between trainee and experimenter GUIs and the scenarios was achieved using a client-server relationship between the two machines implemented by communication sockets. The experimenter's GUI had all the capabilities of the trainee's GUI so that appropriate adjustments, prompting, and demonstrating could be accomplished.

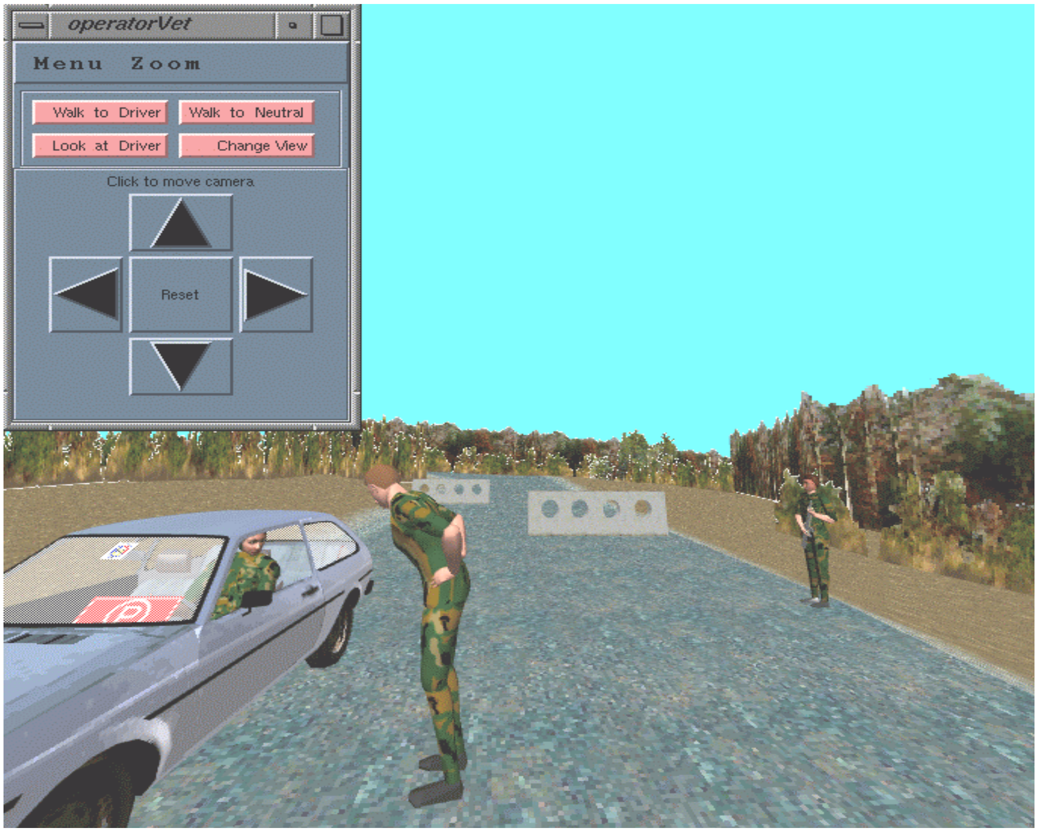


Figure 2. Example of the GUI enabling user navigation in the desktop system.

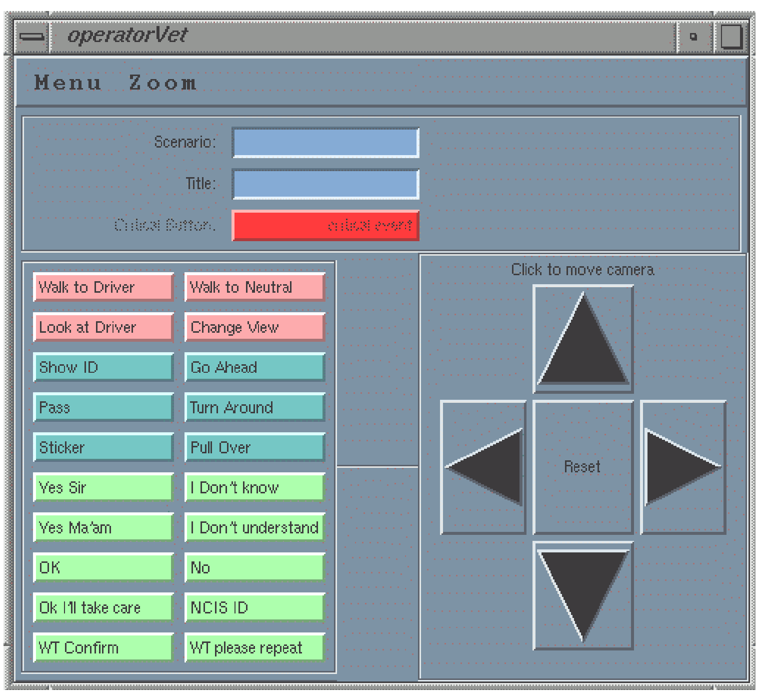


Figure 3. Example of the experimenter's GUI used to intervene and maintain flow of action.

Experimental Method

Fourteen undergraduate students from Old Dominion University with normal or corrected-to-normal vision participated in the study. The task and experimental design were identical to those used in Experiment 1 with a few exceptions. First, participants in the desktop experiments were not administered a simulator sickness questionnaire. Second, in order to decrease the preparation period, the participants were not shown the briefing video. Instead, the critical information from that video was added to their prebriefing package. Third, the sessions were not videotaped. Instead, the sessions were run in adjoining sound attenuated chambers. The participants sat at a desk in one room and the experimenter sat a control panel in the next room. A window between the two rooms allowed the experimenter to "look over the shoulder" of the participant and keep a log of their performance. This log was then used for the AAR for those who participated in two sessions.

Results

The analyses for the desktop experiments were handled differently because there were fewer total participants. Performance was assessed by the total number of errors made by each participant on only the critical scenarios. This approach was taken because so few errors were made on the neutral scenarios. A one-tailed t-test indicated that participants in Group 1 made significantly more errors in their first session (M = 7.43, SD = 3.26) as compared to their second session (M = 3, SD = 2.71), $t(6) = 2.8$, $p < .025$.
The data from participants in Group 2 were compared to those of Group 1 from their second session. As hypothesized, the participants in Group 1 made fewer errors (M = 3, SD = 2.71) than those in Group 2 (M = 8.14, SD = 5.24). A one-tailed t-test indicated that this difference was significant, $t(6) = 2.31$, $p < .05$.

The major findings from the immersive and desktop studies are reproduced in Table 3. The table shows the mean number of errors for only critical scenarios. Due to the large difference in the number of participants in each study, assumptions for statistical analyses are unlikely to be met and therefore no statistical comparison was performed. However, an informal comparison between the two studies shows that the overall pattern of results was consistent. Those individuals who participated in two sessions showed marked improvement in their second session and performed better than another group of individuals who only participated in a single session. More important, however, the overall performance levels for all participants in all conditions were better with the immersive environment than with the desktop system.

Table 3.
Mean Number of Errors for Critical Scenarios in the Immersive and Desktop Studies.

	Group 1	Group 2	Group 1	Group 2
Session 1	6.36 (6.0)	5.47 (3.41)	7.43 (3.26)	8.14 (5.24)
Session 2	1.87 (4.51)		3.0 (2.71)	

Standard deviations appear in parentheses.

Discussion

The primary goal of the present set of studies was to evaluate virtual environment technology as a training tool for military checkpoint duty. It was expected that if participants learned from the experience and feedback they obtained in their first session, they would commit fewer errors in a subsequent session. The results from both sets of experiments support this idea. Participants, on average, made about 60% fewer errors on their second session. These findings clearly show that individuals from an undergraduate college population, with little or no military experience, were capable of learning the fundamentals of performing checkpoint duty in an experiential context. Moreover, these findings are not likely the result of the specific scenarios chosen for study in each session. In both experiments, the Group 2 participants who performed only one session made over twice as many errors on the identical scenarios that Group 1 performed in their second session. Moreover, the performance of the Group 2 participants was also similar to that of the Group 1 participants on their first session.

It could be argued that some of the improvement observed for Group 1 across sessions might be attributable to increased familiarity with the task and procedures over the course of their initial session. If that were true, one might expect these individuals to also show an improvement in performance within their initial session. An analysis of performance on scenarios from the first and second halves of the initial session for these participants, however, revealed no differences. Thus, the performance improvements of the Group 1 participants in their second session were most likely due to knowledge acquired through training and the feedback provided during their AAR.

The pattern of results obtained in the immersive VE and VE desktop studies were consistent. Those individuals who received two sessions performed significantly better on their second session and better than another group who performed only a single session. However, a comparison between studies revealed one important difference. The overall level of errors was higher with the desktop system than with the immersive VE system. Moreover, this difference was observed with both groups in all sessions. Thus, although participants were able to learn effectively with both platforms, the overall levels of performance were better in the immersive VE.

The objective data indicate that the participants responded well to the VE. Most participants were initially unfamiliar with virtual environment technology, yet they acclimated quickly to the environment, became accustomed to the methods of interaction, and interacted with virtual objects rather naturally. On a more subjective level, there was evidence that suggested the participants were "immersed" in the task. Some individuals were observed using hand gestures to motion cars to pull up to the gate and others reached out to try and hold the ID card presented by the driver.

Despite these encouraging results, there were several technological problems with the system that affected how the participants interacted. For instance, the voice recognition/natural language interface was a source of many problems. The voice recognition software used in the present study could not recognize various voice tones, inflections, and accents equally among all participants. Consequently, participants often had to repeat commands. Further, it was apparent that participants, especially when well immersed in the environment and task, added extra words and conversational components to their interactions. Unfortunately, the additional utterances were sometimes misinterpreted by the speech recognition software, which in turn, executed unintended commands. Although a set of responses was preprogrammed to address several categories of unintended user dialogue, the response set was very limited and was insufficient for many utterances. Other participants indicated that they were aware they were dealing with a computer and used less natural speech during their interactions. Occasionally, some actions were not attempted because participants believed the system might not respond appropriately.

The limited fidelity of the vehicle models and virtual humans also had an impact on performance. For example, in the back area of one of the vehicles (the jeep model) there was a box-shaped wheel well that was often misconstrued as a suspicious package. This led some participants to question drivers in what were intended to be neutral scenarios. Also, the Jack agent used in the study was originally developed as an anthropometric/ergonomic model and was not designed for the subtleties of human expression or behavior needed to address a wider range of training objectives important for checkpoint duty. For instance, the poor quality of facial expressions and behaviors generated some ambiguity as what constituted "suspicious" behavior in the Jack agents. During the AAR sessions, participants had to be instructed to adjust their criteria to match the lower fidelity of the agents.

Conclusion

The results from the present studies indicate that individuals can benefit from training in a VE that places greater emphasis on social interaction skills. Individuals with little or no military training were able to learn some of the fundamentals for performing checkpoint duty in an experiential context. These findings should encourage those in the development community to continue to improve and refine the technology required for this class of VEs. Additional work on modeling body gestures, facial expressions, and voice recognition in real-time simulations is needed to develop training for more complex social interactions.

The results from the present study also showed that overall levels of performance were higher with the immersive VE as compared top the VE desktop system. Although this difference was found across groups and conditions, the magnitude of the difference was not dramatic. Thus, the ability to port a similar training experience to a less expensive PC platform without major performance differences underscores the potential for providing greater access to this type of VE training in a much more cost effective medium.

References

Badler N.I., Bindiganavale, R., Allbeck, J., Schuler, W., Zhao, L., & Palmer, M. (2000). Parameterized action representation for virtual human agents. In J. Cassell (Ed.), *Embodied conversational agents* (pp. 256-284). Boston: MIT Press.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*, 203-220.

Khan, T. (2002). Approaching high fidelity human tracking and representation in interactive virtual environments. Masters Thesis. Electrical and Computer Engineering, Old Dominion University.

Knerr, B. W., Breaux, R., Goldberg, S. L., & Thurman, R. A. (2002). National Defense. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation and applications* (pp. 857-872). Mahwah, NJ: Erlbaum.

Loftin, R. B., McKenzie, F., Scerbo, M. W., Badler, N., Phillips, M. A., Garcia, H., Lusso, A., & Seevinck, J. *Virtual Environments for Training: Final Report*. Prepared by Old Dominion University for University of Houston and Office of Naval Research (ONR) under Grant N00014-95-1-1044. December 30, 2002.

Pew, R. W. & Mavor, A. S. (Eds.) (1998). *Modeling human and organizational behavior: Applications to military simulations.* Washington D.C.: National Academy Press.

Interactive Multisensor Analysis Training

Wallace H. Wulfeck II	Sandra K. Wetzel-Smith	Janet L. Dickieson
Space and Naval Warfare Systems Center, Code 2302	Space and Naval Warfare Systems Center, Code 2302	Office of Naval Research Code 34 CM FNC
53560 Hull St. Bldg 586	53560 Hull St. Bldg 586	800 N. Quincy St.
San Diego, CA 92152 USA	San Diego, CA 92152 USA	Arlington, VA 22217 USA

Summary:

The Interactive Multisensor Analysis Training (IMAT) project is aimed at improving the preparation of operational users of undersea-warfare sensor systems. The effort has focused on training at all levels from initial individual training ashore through team, platform, and collective training at-sea, at all skill levels from apprentice sensor operators to senior tactical commanders. Operators and tacticians at all levels need a deep and scientifically accurate, but not necessarily formal, understanding of the physical principles that underlie tactical employment of their sensors. IMAT systems use model-based scientific visualizations, including three-dimensional graphics and animations, to illustrate complex physical interactions in mission-relevant contexts, and to provide interactive virtual laboratories in which the principles can be explored. Concepts in instruction include radiated acoustic characteristics, propagation in range-dependent environments, and sensor properties. Training systems provide exploratory environments in which operators and tacticians can examine the effects of change in any of the variables involved in the end-to-end sequence of emission, transmission, reflection and detection. Sensor settings, environmental conditions, and target characteristics can all be modified through a "what-if" simulation approach. These technologies have been applied effectively in basic and advanced sensor operations/employment courses; in individual and team training simulators, and in on-board training. At the battle-group and theater level, new-technology systems are used for decision support during at-sea exercises and operations, and for post-event reconstruction and performance analysis. This paper describes the IMAT training philosophy and approach, the design of training systems, and training effectiveness.

Introduction

The Interactive Multisensor Analysis Training (IMAT) program is a major visualization effort in the US Navy conducted over the past nine years. The objective in this program is to provide performance support and training systems for extremely difficult cognitive tasks involved in Anti-Submarine Warfare (ASW). The effort has focused on training at all levels from initial training ashore through team, platform, and collective training at-sea, at all skill levels from apprentice sensor operators to senior tactical commanders.

ASW involves the use of a variety of **sensor** systems to locate an opposing **threat** or target submarine in the ocean **environment**, often while avoiding counterdetection**.** In particular:

> ***Threat*** energy ***Sources*** refer to the acoustic, electromagnetic, or electro-optical emission and/or reflection properties associated with undersea objects. This includes the types and parameters of signals emitted by them, such as frequency and amplitude. It also includes the azimuthal or "aspect dependent" variability associated with signals in the case of "passive" systems, and/or the reflective properties and aspect dependencies in the case of "active" systems. A source that is sufficiently energetic to be remotely detectable constitutes a *vulnerability*.
>
> ***Environment*** refers to the transmissive properties of the media through which (acoustic or electromagnetic) energy is propagated. The environment affects the path which energy takes as it is

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

reflected and refracted, and the amount of loss that occurs as a result of transmission through the environment. In the passive case, one-way paths and path loss are studied, while in the active case both outgoing and reflected transmission is involved. Environmental properties that affect transmission include the refractivity of the medium, ambient or other interfering noise in the environment that affects signal-to-noise ratio, directional properties of noise, ocean bottom or terrain absorption and contour, reverberation, and other factors. Path loss is also a function of frequency (and in the case of electromagnetic energy so is refraction).

The ***Sensor*** or energy ***Receiver*** and its associated signal ***Processor*** determine how received vulnerabilities, as affected by the environment, are sensed, manipulated and displayed for human operators, tacticians, and strategists. These systems also are controlled by their operators who are seeking to optimize processing so that highest sensitivity is achieved.

These individual topics cannot be understood in isolation from one another. For example, sensor optimization depends on knowledge of the threat vulnerabilities and the intervening environment. A fundamental training problem is that changes in any of the source, environment, or sensor variables interact with all the others. And, for multiplatform operations (those involving multiple ships and aircraft), these triangular interactions increase exponentially.

ASW is an incredibly complicated task: At all levels from individual sensor operator to senior commander, mission success depends on:

- Correctly anticipating the relative merits of position, speed, maneuver and sensor/weapon employment….
- For each platform in the battle force...
- Against the possible range of threat options…
- Within highly variable environmental conditions, …
- In a rapidly changing combat situation in which intelligent opponents try to avoid detection, confuse identification and gain tactical advantage...
- Where other missions, such as air strike or missile defense, may have equal or greater importance to battle force survival.

To cope with tasks of this complexity, operators and tacticians need a deep understanding of the principles underlying operation and employment of their sensors.

Cognitive Complexity

Complex tasks like those involved in ASW are known to be difficult to learn. Nearly 50 years ago, Piaget noted the difficulty of tasks that require coordinating more than one dimension of variation (Inhelder & Piaget, 1958). Over 30 years ago, the state of Massachusetts (and since then many other governments) found it necessary to impose a unit pricing law, because of the difficulty most people have in determining best value among differently sized and priced containers of the same product—a relatively straightforward two-dimensional task. Feltovich and his colleagues (e.g. Feltovich, Spiro, & Coulsen, 1991) have described some features of tasks and/or problems that make them difficult. We have added a few additional criteria in the list below:

- **Abstract** (versus concrete). Physical phenomena are invisible and the underlying cause and effect relationships cannot be observed. Examples include the propagation of sound, or the patterns of sensitivity of an acoustic sensor.

- **Multi-variate** (versus univariate): Multiple underlying causes can affect an outcome. For example, the refraction of sound in water results from sound-speed variation with depth, which in turn depends on variations among salinity, pressure, and temperature with depth.
- **Interactive** (versus separable or additive). Underlying causes may interact with each other, with outcomes dependent on the interaction of variables in addition to each variable acting separately.
- **Continuous** (versus discrete). The dimensions of variation are continuous For example, speed, pressure, and temperature are all continuous variables. Rather than merely memorizing discrete state changes, the learner must understand the effects of continuous change.
- **Non-linear** (versus linear). The relationship of outcome to an underlying dimension is not a simple straight-line function; rather relationships may be exponential, logarithmic, or even more complex. For example, energy loss in propagation often involves an inverse-square relationship.
- **Dynamic** (versus static). The process of variation itself is the subject of analysis, rather than end or intermediate states. A few frozen moments in time are not sufficient to characterize the underlying variation.
- **Simultaneous** (versus sequential). Outcomes vary continuously with changes in underlying variables, rather than as a succession of states.
- **Conditional** (versus universal). Relationships among variables and outcomes may depend on particular boundary conditions or other contextual events. There may be exceptions to general rules or they may apply only in certain circumstances and not in others.
- **Uncertain** (versus certain). Exact values on underlying variables may not be known precisely; instead they may be interpolations, estimates, or approximations.
- **Ambiguous** (versus unique). The same combination of circumstances may result in multiple outcomes, or the same outcome may be the result of different combinations of circumstances.

Antisubmarine Warfare tasks involve all of these attributes. For example, understanding probability of detection involves the interaction of the target's radiated signals (which vary in three spatial dimensions around the target by frequency by time); the intervening environment which may distort or differentially enhance or attenuate signals at particular frequencies; and detectability of signals by a particular shipboard sensor (with directional sensitivity which also varies in frequency and in three spatial dimensions). This problem is further complicated when such factors as radiated noise variation with target speed and depth, ownship motion effects on directional frequency response of sensors, relative motion between target and sensor, and multipath interactions are considered. There are multiple, continuous, interactive, nonlinear, highly-dynamic dimensions of variability, with uncertainty and ambiguity, throughout the ASW problem space.

Interactive Multisensor Analysis Training (IMAT)

IMAT is designed to make difficult scientific and technical concepts underlying sensor employment and tactical planning comprehensible to their operational users. Operators, tacticians and senior decision makers each need to acquire a deep and scientifically accurate, but not necessarily formal, understanding of the physical principles that underlie tactical sensor employment and planning. To meet this requirement, the IMAT program is

- developing training systems which integrate computer models of physical phenomena with scientific visualization technologies to demonstrate the interactive relationships of threat, environment, and sensor for operator training, and interactions of multiple sensor systems for tactician training from individual platform to battlegroup to theater;

- developing training and performance support systems using modeling and visualization technologies; and
- integrating curricula to provide training on sensor employment and high-level tactical planning skills.

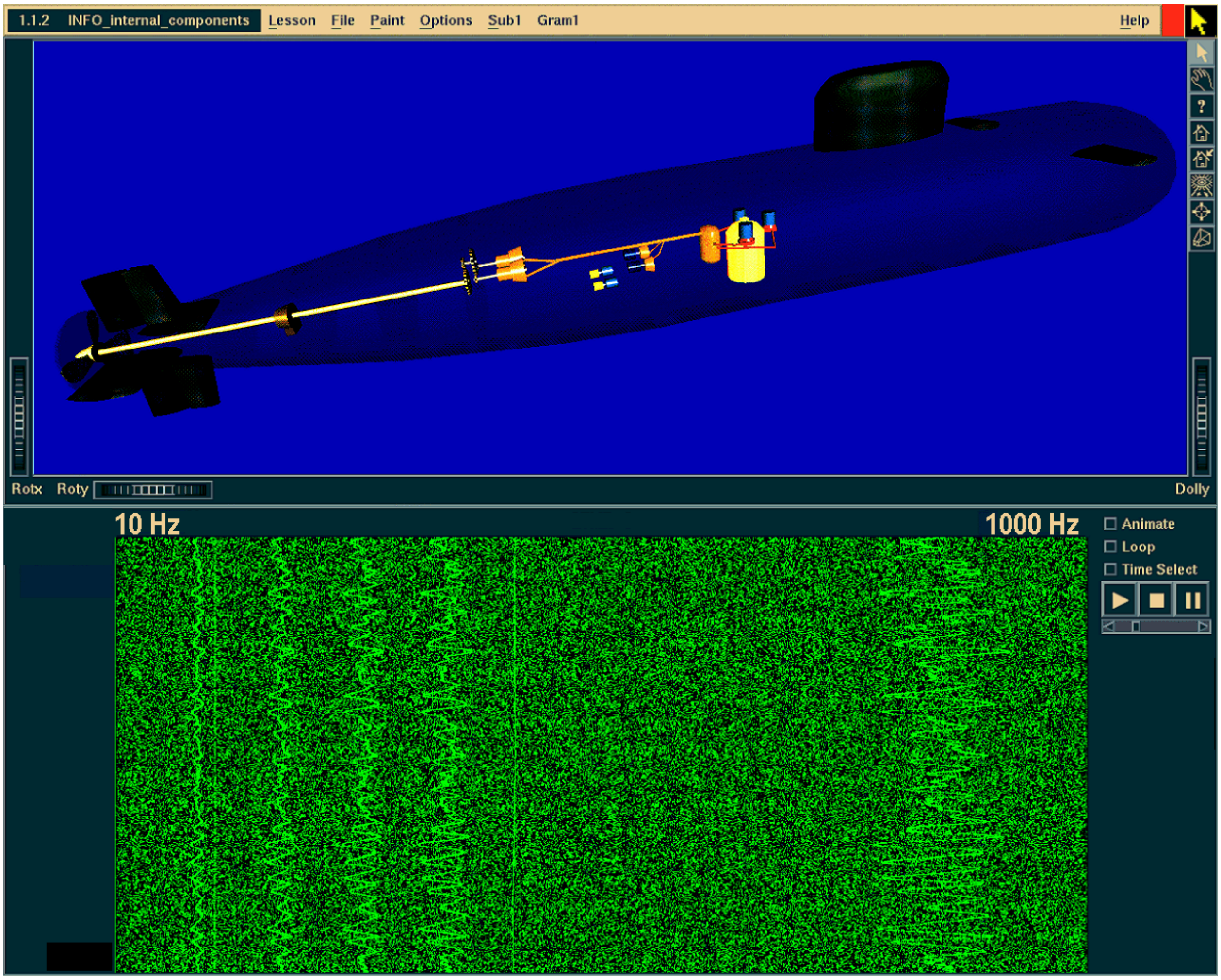
The most important design strategy in IMAT is to use computational physics models as the basic representation of task content. Models of physical phenomena and databases of environmental observations are integrated with scientific visualization technologies to build systems with which users can interact. These systems are used for analysis of cognitive requirements underlying successful task performance, in two ways. First, the physical relationships underlying sensor employment are explicitly identified, and second, task experts can interact with the systems to construct detailed representations of threat-sensor-environment variables and relationships. The visualizations can then be used in training episodes that demonstrate the interactive relationships among all the variables involved. Next, the modeling and visualizations are used during exercises and deployments to identify decision requirements, develop and refine user interfaces, and build tactical decision aids. These are iteratively refined in a build-test-build approach. Ultimately, systems are evolved which provide tactical decision aiding, while at the same time allowing drill-down inspection for tactical situation analysis, as well as exploration of variability in (a) assumptions about presumed target behavior and vulnerability; (b) own-sensor employment; and, (c) environmental effects on sensor effectiveness.

Model-based visualization:

Scientific Visualization is a critically important enabling technology for the design of modern mechanical systems, in medical imaging, and in most of the physical and engineering sciences. There are thousands of university and industrial projects in which visualization is employed to solve engineering, medical, or design projects. Most efforts are aimed at users with graduate degrees in a scientific field, and only a few involve training applications. A recent overview of information visualization is given by Card, Mackinlay and Shneiderman (1999). (Note: An annotated bibliography of Scientific Visualization web sites is located at http://www.nas.nasa.gov/Groups/VisTech/visWeblets.html).

IMAT uses extensive scientific visualization and takes advantage of work in a large number of university and US Navy research laboratories to develop models and databases. These include radiated noise models and databases which describe characteristics of sound sources; oceanographic models and databases which provide high-resolution bathymetric and bathythermographic information, ambient noise, bottom composition, meteorological and other physical effects on propagation; and sensor performance models including recent developments which take account of practical (but tactically relevant) effects such as array motion. In general, the approach has been to adopt these from their controlling organizations who are responsible for validation and verification. IMAT provides the visualization tools that allow the operational and tactical implications of interationships to become observable rather than invisible.

Sample IMAT visualizations are given in figures 1 through 3.



The Acoustic Source Display (Figure 1) is an example visualization of the internal components of a submarine. This is one view into a modeled acoustic laboratory for sound sources. The propulsion and auxiliary systems depicted in the diagram can all be animated to show how they operate, and the animations can be linked to recordings of acoustic data or to an audio simulation to show how acoustic parameters such as frequency are related to the physical operation. The user can select a motor, pump or other object in the diagram, which will highlight it and display a textual description of that object. Frequency lines associated with the component are highlighted on the sound spectrogram in the bottom part of the display. Each object is also linked to a more detailed three-dimensional representation that enables the student to gain a better understanding of how complex assemblies work, why they generate certain signals, and how signals relate to operating mode and speed. Modeled objects in the acoustic laboratory include examples of diesel engines, turbines, reduction gears, pumps, propellers, motors, generators, compressors, and blowers. Each object is also linked to a high fidelity acoustic simulator in which parameters that control the simulation can be varied and explored for instructional purposes. Motors, pumps, clutches, and other components can be activated and deactivated to show variations in operating mode; features such as number of cylinders, power cycles, or gear ratios can be changed; and depth and speed changes can be made, with all changes properly reflected in the visual and auditory displays.

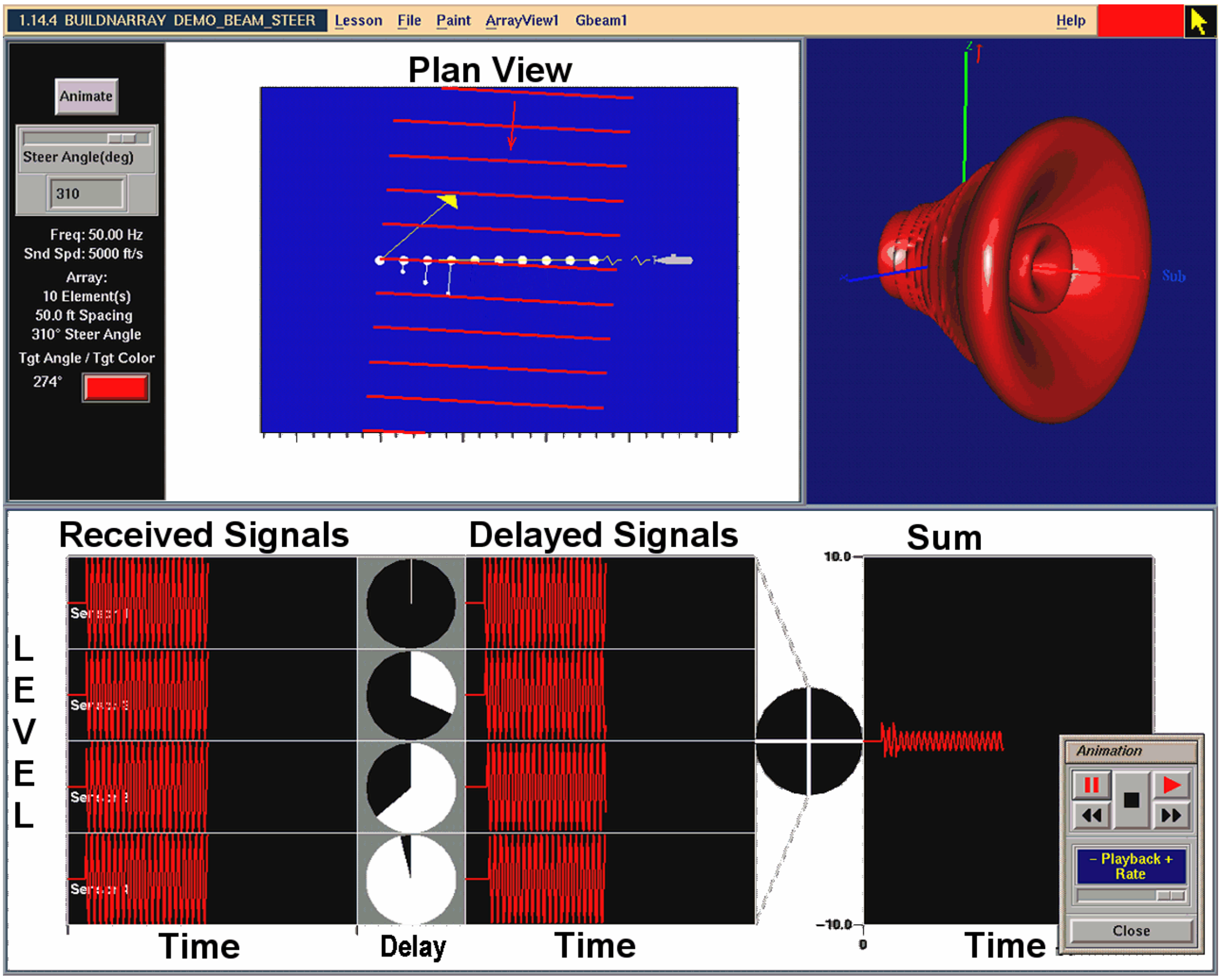


Figure 2 is a view of an interactive model-based animation for explaining the principles of phased array beamforming. Initially, basic concepts, such as element spacing related to frequency, and in-phase vs. out-of-phase arrival, are introduced using a simple two-element array. Later more complicated arrays of acoustic sensors are introduced. In Figure 2, the top right panel shows a 3-D rendered view of the 3dB-down isosensitivity surface for sound of a given frequency arriving at a notional 10-element line array. In the simulation, inter-element spacing and phase delays are all adjustable. The top-center panel shows delays applied to steer sensitivity for a particular frequency in the direction shown by the yellow arrow. "Ghost-elements" corresponding to signal delays are also shown. The bottom panel shows signals arriving at four of the elements, the amounts of delay on those elements, the resulting delayed signals (or the signals arriving at the ghost elements), and their sum. In this case, noise arriving from a direction other than the steer angle is added out of phase and effectively nulled. This interactive simulation is also capable of modeling multi-aperture and other array geometries. The student or instructor can change any of the parameters in these displays, in order to investigate beam width and directivity as a function of array design and employment.

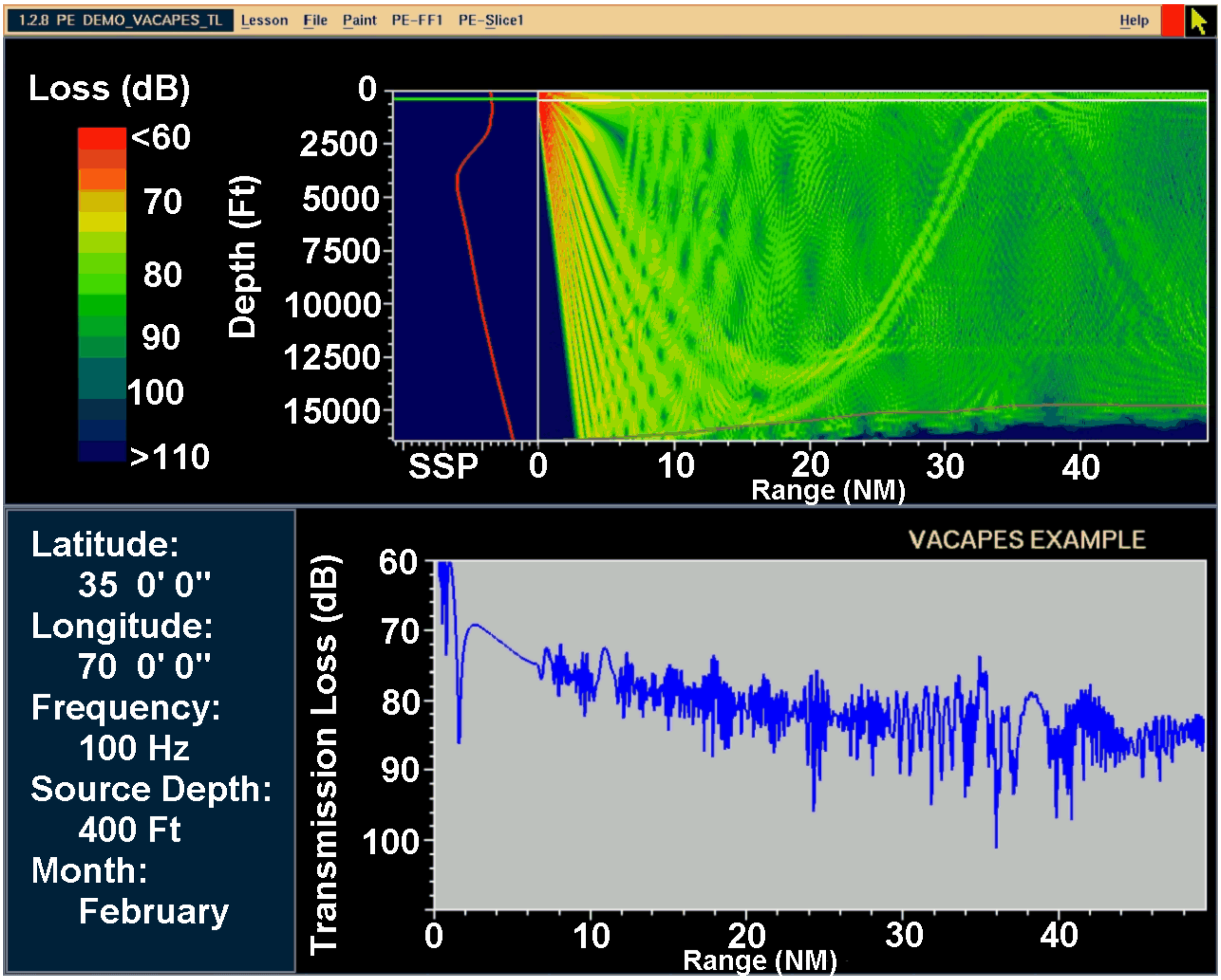


Figure 3 shows a view from the interactive modeling facility for transmission loss in the ocean environment. This allows a visual exploration of sound propagation paths due to reflection and refraction. A sound speed profile (SSP) is displayed on the left of the display. The bottom type, SSP, and bottom contour data can be manually entered or extracted from high-resolution databases. The top right panel shows an example full-field plot of energy loss, with the bottom panel showing transmission loss over range at the sensor depth indicated by the white horizontal line in the top panel. The user can simply drag the depth line to update the transmission loss plot. All the factors that affect transmission loss, such as spreading, absorption by the bottom, and scattering at the bottom and surface, are modeled and contribute to the interactive displays. IMAT includes extensive range-dependent propagation loss models, and databases of environmental data (such as sound speed or bottom absorption) approved by the Oceanographer of the Navy. With these modules, a user can select any geographic location and time of year, extract, view, enter, and/or modify environmental data, specify source and target depths and frequency of interest, and then investigate propagation loss as a function of depth, distance and azimuth from either a sensor or threat.

Task Analysis and development of Training Objectives:

The model-based scientific visualizations in IMAT have also enabled a new approach to the specification of training objectives. The analysis of complex conceptual tasks for training and educational purposes has been a central problem in instructional psychology for at least the past 50 years. The traditional method for analyzing a task, as described in most military training development guidance documents (e.g. NAVEDTRA 130 series in the US Navy), is to identify the components of a task by hierarchically decomposing it into subtasks, skills, and knowledge. Training is then based on these units, and they are tested mostly individually. This unfortunately results in a focus on low-level detail in training, so that students are presented materials as independent topics taught in a serial fashion with limited cause-and-effect explanation as to how those topics interrelate. This approach often leads to instruction in which students read and listen to lengthy descriptions of complex phenomena and memorize large amounts of factual data without any contextual reference. These methods produce graduates who can answer specific factual questions based on memorized information and who can perform procedures but cannot apply that knowledge in operational situations. For example, before IMAT, acoustic operators received instruction about oceanography and the physics of sound as if they were independent of a target submarine's operating environment, mission dependent acoustic characteristics, and the displayed frequency-based information resulting from sensor system engineering design. Operators trained with this approach could recite from memory specific threat emitter parameters, but did not necessarily know how to optimally employ a sensor system against that threat under specific environmental conditions. These operators could answer knowledge-based questions and perform procedures, but when confronted with a variety of real-world situations, they were not able to perform adequately. The operational problems that result include inefficient threat detection and ineffective environmental analysis.

These sorts of performance deficiencies are the result of the application of traditional analysis and instructional methods to these types of tasks. As Feltovich *et al* (1991) point out:

> "...common strategies of simplification...such as teaching topics in isolation from related ones (compartmentalizing knowledge), presenting only clear instances (and not the many pertinent exceptions), and requiring only reproductive memory in assessment are often in conflict with the realities of advanced learning—where components of knowledge are fundamentally interrelated, where context-dependent exceptions pervade, and where the ability to respond flexibly to "messy" application situations is required."

Further, when task analysis results in the introductory instruction for complex interrelated tasks being taught as a series of isolated topics, there may be a detrimental effect on future learning:

> "We have found these discrepancies between introductory and advanced learning often result in situations where the groundwork set down in introductory learning actually *interferes* with successful advanced learning."

More modern conceptual analysis methods take a different approach; they focus on the interrelationships of concepts in a technical domain, and for instructional design purposes they attempt to analyze the thought processes of the performer during performance of a task. However, proponents of cognitive task analysis have not yet had much success at developing their methods so that they can be routinely applied in complex warfighting tasks, and in some cases this approach has led to overly narrow characterizations of Navy training requirements.

The IMAT project has led to a process for conducting conceptual analyses, which involves the following general steps:

a. Define the most complex performance problem for which a training solution is required

b. Identify and refine the variables, and dimensions along which they vary, necessary to model the problem.

c. Obtain or develop mathematical and/or qualitative-process models which relate these variables/dimensions and specify how they interact.

d. Design an interface and display system which facilitates understanding of the variables and their relationships

e. Identify problem scenarios (cases) using the resulting simulation.

f. Validate the problem scenarios by working through them with operators and tacticians.

In general, the process of constructing and validating model-based visualization systems identifies the underlying critical variables, their relationships, and their tactical implications. These then become the enabling concepts and tasks in the analysis. This analytic methodology has now been successfully applied for acoustic, electromagnetic, and electro-optical systems, including revealing employment-training requirements for developmental systems still in test and evaluation.

IMAT Training Techniques

The IMAT effort provided a unique opportunity to integrate and jointly evaluate several of the developing cognitive techniques, including cognitive modeling, situated learning, elaborated explanations, and graphical techniques to promote visualization. The IMAT effort has adopted and tested several modern approaches to complex skill instruction, including the following:

a. *Contextualized/Anchored/Situated Instruction:* Task or job oriented instruction has been found to be more effective in learning, retention, and performance than topic oriented instruction (Semb & Ellis, 1994; Johnson, 1951; Duffy and Jonassen, 1991; Shoemaker, 1960; Steinemann, Harrigan, & VanMatre, 1967; Cognition & Technology Group at Vanderbilt, 1990; Collins, Brown, & Newman, 1989). IMAT has employed this approach in basic and advanced sensor employment and mission planning courses. (Czech, C., Walker, D., Tarker, B., & Ellis J.A., 1998).

b. *Graphic Displays/Interfaces Illustrate Cause and Effect Relationships and Help Concretize Invisible Phenomena and Events:* Research on learning from text has shown that adding pictures or graphics aids learning and retention if they supplement the text in some meaningful way (Dwyer, 1972; Gropper, 1966; Royer & Cable, 1976). Levie and Lentz (1982) in a meta-analysis of illustrated text studies concluded that learning and retention is facilitated by illustrations, if the illustrations are directly related to the text. The IMAT effort has shown that delivering instruction via graphical interfaces to conceptual models has a great effect on subsequent performance, both for apprentice and advanced tactical planning tasks (Wetzel-Smith, S.K. & C. Czech. (1996).

c. *Elaborated Explanations of Complex Tasks and Phenomena*: Providing students with elaborated explanations, analogies, etc. about how and why systems, events, and phenomena are structured and function should facilitate learning and retention. Research on learning skills and learning from text has shown that elaborated explanations enhance the students' mental models and increase retention (Mayer, 1989; Konoske & Ellis, 1991; Smith & Goodman, 1982; Swezey, Perez & Allen, 1991). IMAT has shown the same effects with Navy warfighting tasks (Ellis, J.A. Tarker, B., Devlin, S.E. and Wetzel-Smith, S.K. 1997).

d. *Instructional Sequencing:* Mental model development is facilitated by teaching students to reason about events and phenomena that involve several interrelated variables. Earlier research on sequencing showed that with simplified or isolated tasks, different sequences of instructional events made little difference. However, recent research and theory suggests that for complex tasks, sequencing strategies may have significant effects, and these are being observed in IMAT courses (Czech, C., Walker, D., Tarker, B., & Ellis J.A., 1998).

In each of these areas, before IMAT, little experimental work had been done on the extent to which the findings are generalizable to instruction delivered using simulation- and graphical-interface-based training technologies. Furthermore, there are almost no larger efforts that integrate all these approaches into an overall strategy. IMAT has proven the notion that a combination of these approaches will offer a potent learning environment for promoting acquisition of the kinds of complex skills involved in sensor-system operation and tactical planning.

Conceptual Training during Exercise Planning / Execution / Reconstruction

In tactical employment training, tasks involving planning and mission execution for any particular operating environment require that a planner understands how best to optimize the mix of sensor capabilities to detect and prosecute the threat. At the platform level, planners will have to understand which systems and system settings to select to best detect and prosecute an attack while accurately estimating likelihood of counterdetection and potential vulnerability. The tactician will need to predict the environmental effects on each of the sensors to be able to effectively plan and execute the mission. Previously, the training available to tactical planners tended to familiarize them with available environmental products, but did little to teach them how to apply the information in tactical planning and mission execution. IMAT training now allows tacticians to practice this sort of planning and mission execution, and IMAT systems can be used for timely reconstruction of exercises.

Over the past several years, the IMAT approach to tactical training has been explored in connection with at-sea exercises. During an exercise, IMAT researchers work on board with operators and command personnel to provide additional training in tactical use of the ocean in the context of the ship's performance during each watch period. Results typically indicate important performance improvements in the ship's sonar operations. Moreover, these improvements are retained during subsequent readiness evaluations.

This sort of exercise-based training has been repeated with dozens of submarines and surface ships. . To support these exercises, IMAT researchers often spend many weeks developing mission analyses, including oceanographic workups, sensor performance predictions, and counterdetection assessments, as well as displays and visualizations to deliver the training. The development of deployable mission analysis and reconstruction training will require developing authoring and visualization tools which can support much more rapid scenario (re)construction and display. The experience gained in these efforts has led to further definition of the requirements for deployable mission analysis and reconstruction tools

For interactive mission analysis training, novel methods of curriculum design are needed. IMAT training involves extensive use of case- or situation-based analysis/reconstruction training usable in several types of training scenarios:

- Provide a conceptual overview which gives an integrated expert-model based approach to understanding the variables affecting sonar tactics.
- Run pre-built training episodes which contain teaching points / instructional strategies for developing tactical skill.
- Use analysis capabilities of IMAT visualization systems to explore and critique sonar tactical planning for at-sea exercises.
- Use analysis capabilities of the visualization systems during exercise execution to explore what-if options for ownship and target(s). Use analysis / reconstruct capabilities to match predicted conditions with actual, then modify plans / projections.
- Use analysis / reconstruct capabilities for post-exercise sensor / tactics training assessment to provide "lessons-learned" reinforcement.

Most recently, these techniques have been extended to the force level, wherein several ships and aircraft conduct joint ASW operations. These forces are under the command of a senior tactician who, with his staff, prepares coordinated plans for the ASW problem, then monitors the execution of the plans, and ultimately is responsible for reconstruction and feedback to the units and personnel involved. To support ASW at the force level, IMAT researchers have built visualization systems that can be networked among the platforms involved to support collaborative training and mission execution. The US Navy has established IMAT Fleet Training Teams that deliver training at-sea at the platform and force levels.

Evaluations of Training Effectiveness

Measures of Effectiveness for Visualization-based training systems involve several different criteria. Training and performance aiding systems must first capture the complexity of real-world operations. This is assessed through developmental test and sensitivity analyses for performance prediction systems and through continual build-test-build cycles conducted with operational users of the systems under development. (1) Performance prediction systems must properly implement validated physics models and approved databases of input parameters. (2) These systems must properly model all major physical phenomena known to affect sensor / platform performance in real-world operations. (3) Performance prediction and visualization systems must by usable by experienced operators and tacticians so as to provide support through major phases of their tactical tasks, including planning, search, and prosecution. They should add little additional complexity to operator and tactician tasks. IMAT systems are subject to these kinds of criteria—they have been independently tested, and adopted for use as fleet-approved tactical decision aids by the US and other Navies.

Training improvements are assessed through pretest-posttest and training-vs-control evaluations on tests requiring knowledge and skill application on scenario-based tactical reasoning problems. IMAT visualization-based training consistently result in additional operator / tactician knowledge and skill. Evaluations of training effectiveness in ASW schools indicate that IMAT is among the most successful training technologies ever introduced in the US Navy (Committee on Technology for Future Naval Forces, 1997):

- IMAT students outperform students in conventional instruction, and in many cases score higher than qualified fleet personnel with 3 to 10 years experience. Evaluations consistently show gains of two to three standard deviations on comprehension, reasoning, and problem solving tasks. Overall, the IMAT approach is much more effective than conventional lecture instruction, or new technologies such as interactive video or computer-based training.
- Instructors report that IMAT increases their ability to teach difficult topics, respond to student questions, and reinforce critical principles.
- IMAT students score higher on attitude scales measuring attention, relevance, confidence, and satisfaction than students in standard Navy classrooms or students in specially designed individualized computer based training.

Finally, improved training should result in observable improvements in operator and tactician performance during exercises and operations, for example in improvements in the quality and timeliness of tactical plans and decisions made during tactical execution. Measures include: (1) better utilization of platforms (e.g. less overlap in sensor coverage), (2) deeper user examination of alternative courses of action during planning and execution, (3) improved availability and processing of information necessary to support planning and execution monitoring (e.g. minimize time spent collecting information while maximize time available for information analysis).

Ultimately of course, improved training and performance support systems should result in measurable improvements in military capability. Measures should reveal (1) increased detection / engagement ranges, (2) increased search rates or area coverage, (3) more rapid localization and classification, (4) reduction in prosecution of false contacts, (5) reduction in counter-detection and counter-attack vulnerability, and (6) increased tactical control or advantage. All of these measures are highly dependent on the boundary conditions for tactical exercises and operations – they can vary widely depending on any or all of the variables (e.g., threat parameters, environmental conditions, sensor employment, target behavior and tactics, ambient noise, etc.) that normally complicate military operations. IMAT researchers and training-team instructors have provided decision-aiding systems and advanced training to individual ships and to commanders and command staffs for battle groups. Independent evaluations of battle-group performance reveal improvements in all these measures.

Conclusion

The IMAT program is providing training and performance support systems designed to make difficult scientific and technical concepts comprehensible to the operational users of advanced sensor systems. The program is (1) developing systems which integrate computer models of physical phenomena with scientific visualization technologies to demonstrate the interactive relationships of threat, environment, and sensor for operator training, and interactions of multiple sensor systems for tactician training; (2) developing training and performance support systems using modeling and visualization technologies; (3) integrating curricula to provide training on high-level sensor operation and tactical planning skills; and (4) developing modeling and visualization tools for use at sea both for training and as tactical decision aids.

The IMAT vision is to integrate training, operational preparation, tactical execution, and post-mission analysis into a seamless support system for developing and maintaining mission-related critical skills. In many ways, IMAT is a prototype for future human performance support systems that transcend traditional shore school and course structures to span career-long skill development from apprentice to master levels, across missions, platforms, and communities.

Acknowledgements

The Interactive Multisensor Analysis Training (IMAT) program is a joint effort of the Space and Naval Warfare Systems Center, San Diego and the Naval Surface Warfare Center Carderock Division, supported by several contractor companies including Presearch Incorporated and Anteon Corporation. Over 50 people have worked on the IMAT effort, and the authors in particular thank Bill Beatty, Rich Loeffler, Eleanor Holmes, Kent Allen, Dan Lawrence, Jeff Raver, and Rick Krutz for their contributions to the program. Portions of the work described in this paper are supported by the Capable Manpower Future Naval Capability program in the Office of Naval Research, Code 34.

References

Card, S., Mackinlay, J. & Scheiderman, B (Eds) (1999). *Readings in Information Visualization: Using Vision to Think.* San Francisco: Morgan Kauffmann. ACM Press.

Cognition & Technology Group at Vanderbilt (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher, 19(6),* 2-10.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Erlbaum.

Committee on Technology for Future Naval Forces, (1997), National Research Council, *Technology for the United States Navy and Marine Corps, 2000-2035 Becoming a 21st-Century Force: Volume 4: Human Resources,* National Academy Press, ISBN POD519, Also available on line at http://www.nap.edu/catalog/5865.html and http://www.nap.edu/html/tech_21st/hrindex.htm.

Czech, C., Walker, D., Tarker, B., & Ellis J.A., (1998). *The Interactive Multisensor Analysis Training (IMAT) System: An Evaluation of the Airborne Acoustic Mission Course*. NPRDC TR 98-2. San Diego: Navy Personnel Research and Development Center, 1998. DTIC AD Number ADA338076

Duffy, T.M., & Jonassen, D. H. (1991). New implications for instructional technology. *Educational Technology, 31(5),* 7-12.

Dwyer, F. M. (1972). *A guide for improving visualized instruction.* State College, PA: Learning Services.

Ellis, J.A. Tarker, B., Devlin, S.E. and Wetzel-Smith, S.K. (1997) *The Interactive Multisensor Analysis Training (IMAT) System: An Evaluation of Acoustic Analysis Training in the Aviation Antisubmarine Warfare Operator (AW) Class "A" School.* NPRDC-TR-97-3, San Diego: Navy Personnel Research and Development Center, 1997. DTIC AD Number: ADA328827

Ellis, J. A., & Parchman, S. (1994). *The interactive multisensor analysis training (IMAT) system: A formative evaluation in the aviation antisubmarine warfare operator (AW) class "A" school* (Tech. Note NPRDC TN 94-20). San Diego: Navy Personnel Research and Development Center.

Feltovich, P. J., Spiro, R. J. & Coulson, R. L. (1991). Learning, teaching, and testing for complex conceptual understanding. (Conceptual Knowledge Research Project Technical Report No. 6). Springfield, IL: Southern Illinois University School of Medicine. Also in N. Frederiksen, R.J. Mislevey, & I.I. Bejar (Eds.), Test theory for a new generation of tests (pp 181-217). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gropper, G. L. (1966). Learning from visuals: Some behavioral considerations. *AV Communication Review, 14,* 37-70.

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. New York: Basic.

Johnson, H. O. (1951, June). *The development of more effective methods of training electronic technicians.* Washington, DC: Working Group on Human Behavior Under Conditions of Military Service, Research and Development Board, Department of Defense.

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. 3*Educational Communication and Technology Journal, 30,* 195-232.

Konoske, P. J., & Ellis, J. A. (1991). Cognitive factors in learning and retention of procedural tasks. In R. F. Dillon, & J. W. Pellegrino (Eds.), *Instruction: Theoretical and Applied Perspectives.* New York: Praeger.

Mayer, R. E. (1989). Models for understanding. *Review of Educational Research, 59(1),* 43-64.

Semb, G. B., & Ellis, J. A. (1994). Knowledge taught in school: What is remembered? *Review of Educational Research, 64,* 253-286.

Shoemaker, H. (1960). The functional context method of instruction. *IRE Transactions of Education, E-3,* 54-58.

Smith, E., & Goodman, L. (1982). *Understanding instructions: The role of explanatory material.* Cambridge, MA: Bolt Beranek, & Neuman.

Steinemann, J. H., Harrigan, R. J., & VanMatre, N. H. (1967). *A performance-oriented electronics training program; IV. Fleet followup evaluation of graduates of all classes* (Special Report NPRDC SRR-68-10). San Diego: Navy Personnel Research and Development Center.

Swezey, R. W., Perez, R. S., & Allen, J. A. (1991). Effects of instructional strategy and motion presentation conditions on the acquisition and transfer of electro-mechanical troubleshooting skill. *Human Factors, 33,* 309-323.

Wetzel-Smith, S. K., Ellis, J. A., Reynolds, A. M., & Wulfeck W. H. (1995). *The interactive multisensor analysis training (IMAT) system: An evaluation in operator and tactician training.* (Tech. Report NPRDC TR 96-3). San Diego: Navy Personnel Research and Development Center.

Wetzel-Smith, S.K. & C. Czech. (1996). *The Interactive Multisensor Analysis Training System: Using Scientific Visualization to Teach Complex Cognitive Skills."* NPRDC TR 96-9, San Diego: Navy Personnel Research and Development Center, August 1996. DTIC AD Number: ADA313318

Evaluation of Collective Training in a Distributed Simulation Exercise

Ir. Arjan J.J. Lemmers
National Aerospace Laboratory NLR
Anthony Fokkerweg 2
1059 CM AMSTERDAM
THE NETHERLANDS
tel. +31 20 511 3581
e-mail Lemmers@nlr.nl

Bernd Rollesbroich
CAE Elektronik GmbH
Steinfurt 11
52222 STOLBERG
GERMANY
tel. +49 (0) 2402 106 508,
e-mail brolles@cae-gmbh.de

M.Sc. Timo Hartikainen
Instrumentointi Oy
Sarankulmankatu 20
33900 TAMPERE
FINLAND
tel. +358 40 3400 661
e-mail timo.hartikainen@insta.fi

Francisco J. Rodriguez Carvajal
Sainsel Sistemas Navales, SA
Ronda del Tamarguillo, 29
41006 SEVILLA
SPAIN
tel. +34 954 93 53 72,
e-mail Franciscojavier.rodriguez@sainsel.abengoa.com

ABSTRACT

The military interest in Synthetic Environments (SEs) is beginning to change from thinking about the relevance of capability demonstrators to SEs being used to support distributed simulation exercises. Several European nations are actively promoting the use of SEs for Collective Training purposes to increase the military co-operation in Europe. In order to overcome the obstacles for use of distributed simulation exercises across Europe, it is important that a common European process, tools and standards are defined. This paper is focussing on the evaluation issues in distributed simulation exercises. It will outline the process of defining the evaluation needs, the identification of the functional and technical evaluation requirements and the definition of the Common Evaluation Framework (CEF). The CEF comprises processes, models, methods and presentation means for results presentation and distribution. Three supporting prototype tools for the CEF have been developed under the Euclid RTP11.13 programme: the Evaluation Definition Tool (EDT), the Evaluation Definition Selection Tool (EDST) and the Execution Evaluation Tool (EET). The EDT is used by the military user to define evaluation objectives and criteria. The EDST gives possibilities for searching and selecting evaluation objectives from a pool of ready-made objectives. The EET supports the user in post-processing the outputs acquired during the exercise and in the analysis and evaluation of the results and trainees.

INTRODUCTION

It is well known that in the training domain, individual skills are taught much more effectively using tailor-made training devices. The main advantage of SEs is in training certain collective skills. For the purpose of this study, the use of networked simulations in a training centre that satisfy particular collective training objectives is not considered to be an SE; although the individual simulators can be used as assets in an SE. The training SEs considered in this programme are those that require a range of different simulations (either manned or constructive) to be identified, configured and networked. This type of SE

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held in Genoa, Italy, 13-15 October 2003, and published in RTO-MP-HFM-101.

could be used to replace joint and coalition training exercises that are currently conducted using real equipment that is costly to operate.

The military interest in SEs is beginning to change from thinking about the relevance of capability demonstrators to SEs being used to support real programmes. Several European nations are actively promoting the use of SEs within their countries. However, because of the distributed nature of SEs they are also well suited to support the increase in military co-operation in Europe; through coalition training, multinational operations and multinational equipment acquisition. In order to promote the use of distributed simulation exercises across Europe, it is important that a common European process, tools and standards are defined.

Euclid RTP 11.13 is a major European initiative to promote the use of Synthetic Environments (SEs). The title of the programme ‘Realising the Potential of Networked Simulations in Europe’ reflects the fact that although SEs are currently being used to support defence programmes in Europe, their full potential is not currently being realised. The aim of the project is to ‘overcome the obstacles that prevent SEs being exploited in Europe by developing the SE Development Environment (SEDE). SEDE provides a facility that will assist the different types of SE users i.e. Problem setters, Problem Solvers, and SE Implementers, so that SEs can be delivered faster, better and cheaper. It will achieve this by providing a common shared data environment, providing facilities for managing the data generated by an SE project and making information about available SE assets and best practices readily available to SE users. The SEDE comprises of five main components: a Process, Repository, SE Management Tool, SE Tools (both COTS and those being prototyped in Euclid 11.13) and a Knowledge Base.

The Federation Development and Execution Process (FEDEP) has been used as a baseline for the process but has been modified and extended where it has been found to be lacking. One of the shortcomings was that the FEDEP did not cover the complete lifecycle of a SE. Therefore one of the extensions is a new step “Evaluation”. The resulting RTP 11.13 process is known as the Synthetic Environment Development & Exploitation Process (SEDEP). The purpose of the SE tools is to assist the SE users in performing their roles. Several COTS tools are already available for supporting some of the SEDEP activities and additional tools are being prototyped where none currently exist. A key requirement for supporting the SEDEP is the use of common data formats, captured in the SEDE data model. In this way, the data created by one tool can be read by the tool supporting the next activity in the process. All tools will have access to the RTP11.13 Repository, which will provide the mechanism for transferring data between them. The Data Interchange Formats (DIFs) defined by the High Level Architecture (HLA) are being extended to other areas supported by the SEDEP. The SEDE may access the data from a local Repository or from one that has been distributed over a Wide Area Network (WAN).

This paper is focussing on the evaluation issues in the SEDE. It will outline the process of defining the evaluation needs, the identification of the functional and technical evaluation requirements and the definition of the actual Common Evaluation Framework (CEF). The CEF comprises processes, models, methods and presentation means for results presentation and distribution. Three supporting prototype tools for the CEF have been developed under the Euclid RTP11.13 programme: the Evaluation Definition Tool (EDT), the Evaluation Definition Selection Tool (EDST) and the Execution Evaluation Tool (EET). The EDT is used by the military user to define evaluation objectives and criteria. The EDST gives possibilities for searching and selecting evaluation objectives from a pool of ready-made objectives. The EET supports the user in post-processing the outputs acquired during the exercise and in the analysis and evaluation of the results and trainees.

WHAT IS COLLECTIVE TRAINING

Collective training (CT), in military terms, has been defined by the NATO SAS-13 Military Application Study on 'NATO Mission Training via Distributed Simulation' as training which

- "…involves 2 or more 'teams', where each team fulfils different 'roles', training in an environment defined by a common set of collective training objectives (CTOs)".
- A team is defined as "a number of individuals who may have different 'tasks' within that team but whose operational remit is to fulfil a specific role e.g. a tactical 4-ship in a ground-attack role."

CT applies to training of military groups in order to maintain or improve the groups' ability to perform in terms of service. The trained group in a single CT exercise may include multiple crews of similar or dissimilar vehicles and possibly different domains. To improve the fidelity the exercise may also involve Computer Generated Forces (CGF) as in ref. [1], or may be conducted as embedded training. CT is considered as the training required to prepare cohesive teams and units to accomplish their assigned operational missions. CT is part of a continuous process of unit training and is generally conducted within operational units, or specialist training facilities available to operational units on timeshare basis. CT exercises individual tasks, skills and responsibilities and collective command and control responsibilities.

Generally, the teams that participate in a CT exercise comprise a battle group, possibly including complete command and control functions. In addition to rehearsing within one military application, CT is used in joint exercises where the cooperation between e.g. air force, navy and army is practised. From the performance point of view, performance of a team is a product of the competencies of the different individuals. The performances of the individuals affect the performance of the teams and the whole trained battle group. According to ref. [2] competencies may be defined for successful performance. The competencies may be divided into individual, intra-team and inter-team competencies. When these competencies are applied to the performance of certain teams and missions, they become mission essential competencies (MEC). In CT applications the inter-team MECs are considered more important than the individual or intra-team MECs. Although the same underlying skills may be employed at different levels of MECs, such as communication, co-ordination etc., these are applied in a different context (ref. [2], [3]).

In order to measure the effectiveness of the simulation it is essential to be able to assess whether adequate CT is being achieved. However, the effectiveness of CT has been hampered by the difficulties involved in evaluating collective performance and feeding this information back to the personnel being trained. Whilst training objectives at an individual level can be defined in relation to specific tasks and are role specific, it is much more difficult to define a set of measurable collective training objectives, which apply to all members of a CT audience. The distributed, group nature of CT makes any measurement of performance difficult, with the level of difficulty increasing as the size of the unit under training increases. Currently only basic, usually subjective, forms of measurement exist at the sub-unit level and above. The EUCLID RTP11.13 team has identified this fact as a major obstacle for realising the potential of using SEs for CT purposes. The process of setting up SEs for CT needs to be refined further in order to derive more robust CT metrics, which could be used in live or synthetic exercises. To achieve this work should be done in further developing the process and tools to support evaluation matters for CT.

THE MISSING STEP: EVALUATION

The requirements for the process that should be developed in RTP11.13 were:

- Provide support to encourage the use of SE technology on military programmes;
- Provide guidance for SE developers and users to plan and perform the different activities necessary to produce the required products and results;

- Promote good practice for developing SEs on time and within budget;
- Promote reuse of products (federation, federates, components) and results;
- Provide a framework for a tool set to reduce the cost and time for producing and using SEs.

Because the Federation Development and Execution Process (FEDEP) (Ref. [4]) has already been widely adopted and is already supported by several COTS tools, the decision was made in Euclid 11.13 to use the FEDEP as a baseline instead of developing a new SE process from scratch. The FEDEP version 1.5 comprised 6 steps:

- Step 1: Define Federation Objectives. The federation user and federation development team define and agree on a set of objectives and document what must be accomplished to achieve those objectives.
- Step 2: Develop Federation Conceptual Model. Based on the characteristics of the problem space, an appropriate representation of the real world domain is developed.
- Step 3: Design Federation. Federation participants (federates) are determined and required functionalities are allocated to the federates.
- Step 4: Develop Federation. The Federation Object Model (FOM) is developed, federate agreements on consistent databases/algorithms are established, and modifications to federates are implemented (as required).
- Step 5: Integrate and Test Federation. All necessary federation implementation activities are performed, and testing is conducted to ensure that interoperability requirements are being met.
- Step 6: Execute Federation and Prepare Results. The federation is executed, outputs are generated, and results are provided.

The analysis performed in EUCLID 11.13 identified among others the following limitations of the FEDEP:

- FEDEP does not support all good practice management activities.
- FEDEP does not provide assistance to all the different types of SE users.
- FEDEP does not cover the complete lifecycle of an SE, it focuses on the development part and the analysis and evaluation activities are lacking.
- FEDEP does not explicitly identify products at each step of the process.

EUCLID 11.13 has taken the initiative to enhance and extend the FEDEP process where it was perceived to have limitations. The resulting Euclid 11.13 process is known as the Synthetic Environment Development & Exploitation Process (SEDEP) (ref. [5]). The use of the term SEDEP has been chosen to reinforce its close links with the FEDEP whilst promoting its more general use for developing SEs. To overcome the limitation of the FEDEP, in the SEDEP the following enhancements were made:

- The SEDEP is matched with the different recommended activities of the “Capability Maturity Model-Integrated” (CMMI) for good practice development. This makes SEDEP compatible with the Standard System Engineering Process.
- The SEDEP introduces two new steps (see Figure 1) dedicated to
  a) Support the users in determining the suitability of the SE in solving their problem and estimating project parameters such as cost, duration, risk etc.
  b) Analyse the execution outputs and evaluating the results.

- SEDEP provides an overlay representation to allow traceability of specific technical objects or parameters along the full process.
- SEDEP explicitly identifies and defines inputs and outputs for each step and activity of the process.
- SEDEP specifies the use of a repository to provide a means of storing the information about the SE and for support tools to transfer data between the different phases of the process.
- SEDEP explicitly identifies and defines the different library components in the repository used by the different steps and activities.
- SEDEP provides capability for components, specifications and definitions reuse.

It should be noted that the current steps of the FEDEP exist as a sub-set of the SEDEP (see Figure 1). It is intended that long term, the two processes will merge and that there will only be a single SE process.

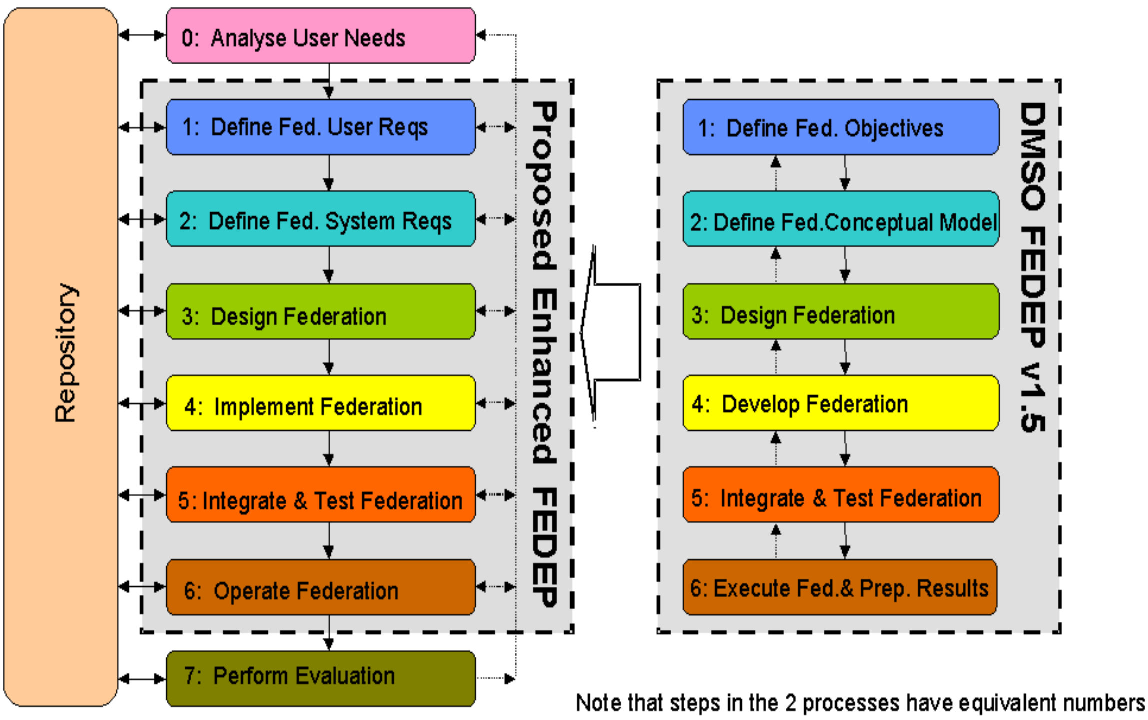
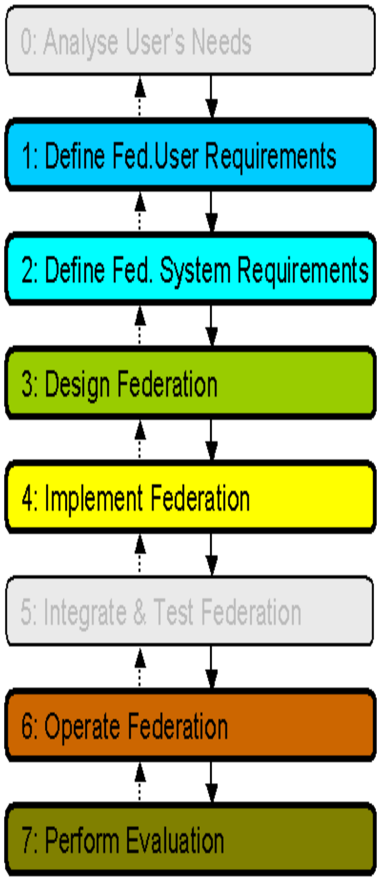


Figure 1: SEDEP Relationship with FEDEP.

The FEDEP developers are currently working on transforming the FEDEP v1.5 into a IEEE standardized process. Discussions took place with the FEDEP Productisation Group to influence the development of the FEDEP before it became an IEEE recommended practice. RTP 11.13 proposed several new steps, activities, and tasks derived from their results in SEDEP v1.0. 16 changes were adopted which are now included in the IEEE 1516.3 FEDEP version. The most important contribution of Euclid RTP 11.13 is the new FEDEP step 7 “Analyse Data and Evaluate Results”.

Because of the continuous discussion about evaluation in the SEDEP and the experience gathered during the tool development it was recognized that there are a lot of activities and tasks relevant for evaluation in different steps of the SEDEP.



In step 1 "Define Federation User Requirements" the Problem Setter and Problem Solver must define which behaviour, skills, characteristics, tactics, procedures, functionality, etc. should be analysed and evaluated (evaluation objectives). The results of step 1 are required to determine the criteria, methods, algorithms, questionnaires, checklists, and presentation information, which have to be used to perform the evaluation. The criteria, methods, and algorithms are defined in detail within several activities in step 2, 3 and 4.

In step 2 "Define Federation System Requirements" the evaluation related information can already be very detailed, but it is also possible that only the name of a new algorithm is defined in this step and that the details are further elaborated in step 3.

In step 3 "Design Federation" the output of the previous step has to be completed and transformed into a generic format. The result of step 3 is a complete and correct mathematical description of all formulas required to post-process the data logged during the execution phase.

In the context of evaluation the purpose of step 4 "Implement Federation" is to produce algorithms and formulas in tool specific formats. The algorithms are allocated to suitable mathematical tools according to the tools' capabilities to apply the required methods. The respective tools are used in step 7 to analyse and evaluate the execution outputs.

In step 6 "Operate Federation" all necessary execution data is collected. This includes filled in questionnaires and checklists. The collected data is filtered and transformed into a generic format. After the execution of the Synthetic Environment these 'prepared execution outputs' are analysed and evaluated in step 7 "Perform Evaluation" to provide the desired feedback.

THE COMMON EVALUATION FRAMEWORK

One of the key factors to reduce the cost and time scale of creating and utilising SEs is the reuse of SE components, specifications, and definitions. To enable and facilitate the reuse of evaluation data it is important to define standards for the type, structure, and format of this data. Euclid RTP 11.13 identified that there is a lack of information about evaluation of SEs. The information available was collected and analysed to provide a baseline for evaluation called the Common Evaluation Framework (CEF). The CEF captures evaluation aspects in general, i.e. regardless whether the SE is used for Collective Training, Mission Rehearsal, or Simulation Based Acquisition.

Before an exercise can be executed data like criteria, methods, and algorithms must be defined. This data must be taken into account for the SE design and implementation in order to meet the evaluation requirements. Within RTP 11.13 the term "Evaluation Need" describes the different items that have to be defined for a complete evaluation (Figure 2).

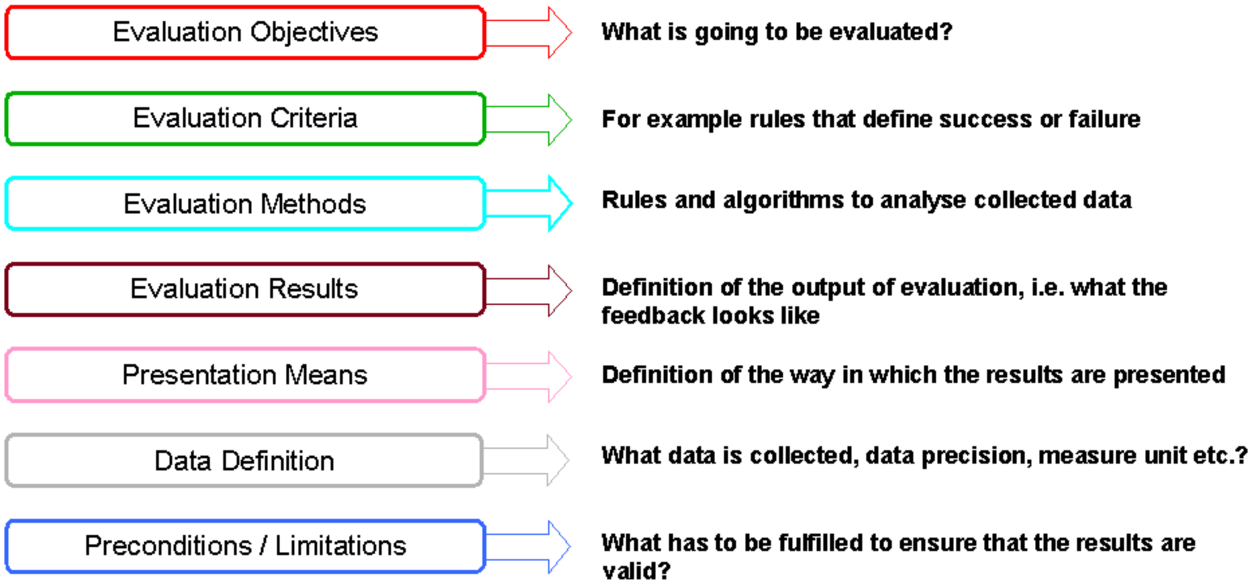


Figure 2: Evaluation Need.

An Evaluation Need consists of Evaluation objectives, criteria, methods and rules, data definition, preconditions, Evaluation data collected during the execution of the SE and the presentation of Evaluation results (i.e. feedback). The Evaluation results are presented with different means (e.g. 3-D visualisation, map drawings, images, text) depending on the purpose and target of the Evaluation. The Evaluator decides the presentation means of the evaluation. Evaluation addresses both the Evaluation of the success or failure of the exercise (presented in military terms) and the assessment of the value of the synthetic environment in satisfying the user's needs in training, rehearsal, acquisition etc.

The Evaluation Objective can be derived from the user's needs to describe a goal of the evaluation in a high-level format. Due to the objective a criterion will be determined that is used to judge the quality of the collected data. In this case quality is related to the performance of the evaluated object and not e.g. to the data completeness. The input data (parameters, variables etc.) that are needed for applying a criterion are produced by applying methods to analyse the collected data. It is not absolutely necessary to process the results of the data analysis within a criterion to receive suitable evaluation results. Applying a criterion or a method or both produces evaluation results. The procedure to generate the evaluation result depends on the evaluation objective. The evaluation results are represented by presentation means, e.g. graphs, charts, or textual reports.

The definition of the data used in criteria and methods contains the name of the data item, precision, unit, etc. The data represented by the elements of an Evaluation Need can be used to determine the Evaluation System Requirements. This is a low-level description containing the technical and functional requirements for the SE that are important to meet the user's expectations of evaluation.

Example of an Evaluation Need:

In the following the concept of the "Evaluation Need" is explained using a simple example from the air force domain. Several fighter aircrafts have to keep the correct flight formation for optimal observation of all sectors around the aircraft. The line abreast formation and its constraints are shown in Figure 3. To simplify the example only the distance between the wingman and the lead and the aircraft's altitude will be analysed.

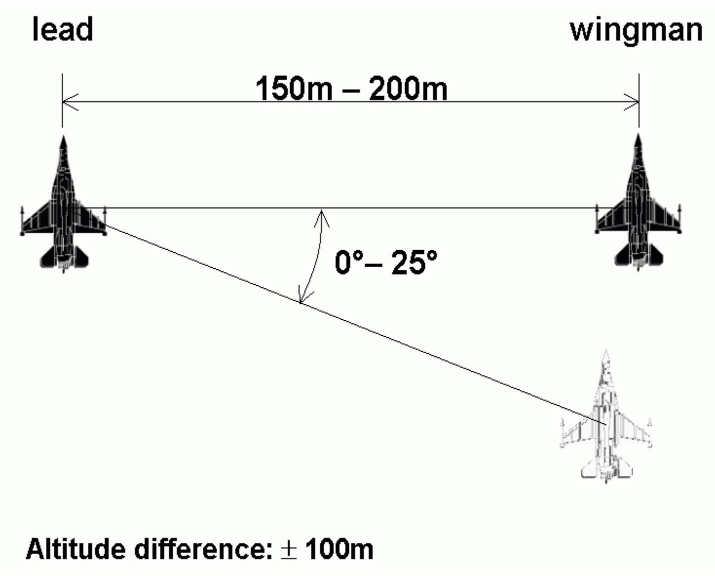


Figure 3: Line Abreast Formation.

The position of aircraft is identified by Lat/Lon coordinates (degree°/minute'/second''). Altitude is provided in feet. Flight time while in line abreast formation is provided in seconds. The distance between flight lead and wingman can be calculated approximately by the following algorithm:

distance = (1,852 x 60 + arc cos (sin P1 x sin P2 + cos P1 x cos P2 x cos LD)) / 1000   unit: m

P1: Latitude Position Flight Lead in decimal degrees

P2: Latitude Position Wingman in decimal degrees

LD: Longitude Difference in decimal degrees

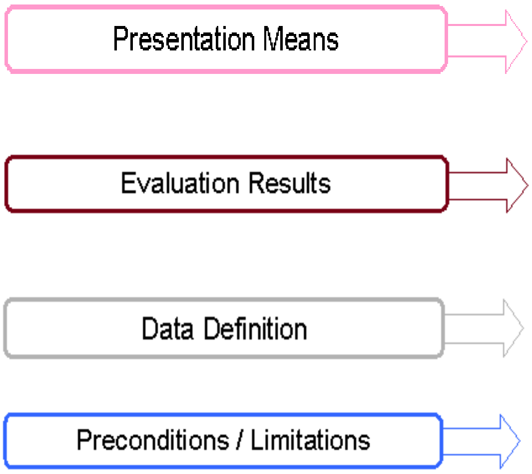
The Longitude and Latitude coordinates are transformed into decimal degrees by applying the following formula:

Decimal degrees = degree° + (minute' + second'' / 60) / 60

To evaluate the altitude the altitude difference must be transformed from feet into meters (1ft = 0,3048m).

To gain a better understanding of the Evaluation Needs elements this information is linked to the different elements:

Evaluation Objectives	Evaluate the line abreast formation correctness.
Evaluation Criteria	If (Lead-Wingman-Distance > 200m) or (Lead-Wingman-Distance < 150m) or (Altitude difference > 100m) or (Altitude difference < -100m) then Formation is not correct.
Evaluation Methods	Lead-Wingman-Distance = (1,852 x 60 + …) /1000 Decimal degrees = degree° + (minute' + second'' / 60) / 60 Altitude difference = (AltLead – AltWingman) / 0,3048



The evaluation report should contain the following fixed sentence: The line abreast formation was … % of the flight time incorrect.

The evaluation result is the percentage of correctness, which is determined by: adding up the time when the rule of the criterion was not true and dividing this time by the complete formation flight time.

The data used in algorithms, methods, and rules must be defined, e.g. Name: AltLead, Type: Integer, Unit: Foot or Name: Flight time, Type: Integer, unit: Seconds.

One limitation is that at low altitude the wingman should not fly lower than the lead.

EVALUATION DATA

A key requirement for supporting the SEDEP is the use of common data formats, captured in the SEDE data model. In this way, the tool supporting the next activity in the process can read the data created by the tool supporting the previous activity. For storing the evaluation related information generated throughout the process, the Evaluation Data Structure was developed. This data structure covers all the evaluation activities in the SEDEP process and it is a part of the SEDE data model. Even though only Collective Training is considered in this paper, the Evaluation Data Structure supports also at least Mission Rehearsal and Simulation Based Acquisition application areas. Figure 4 shows the top-level elements of the Evaluation Data Structure.

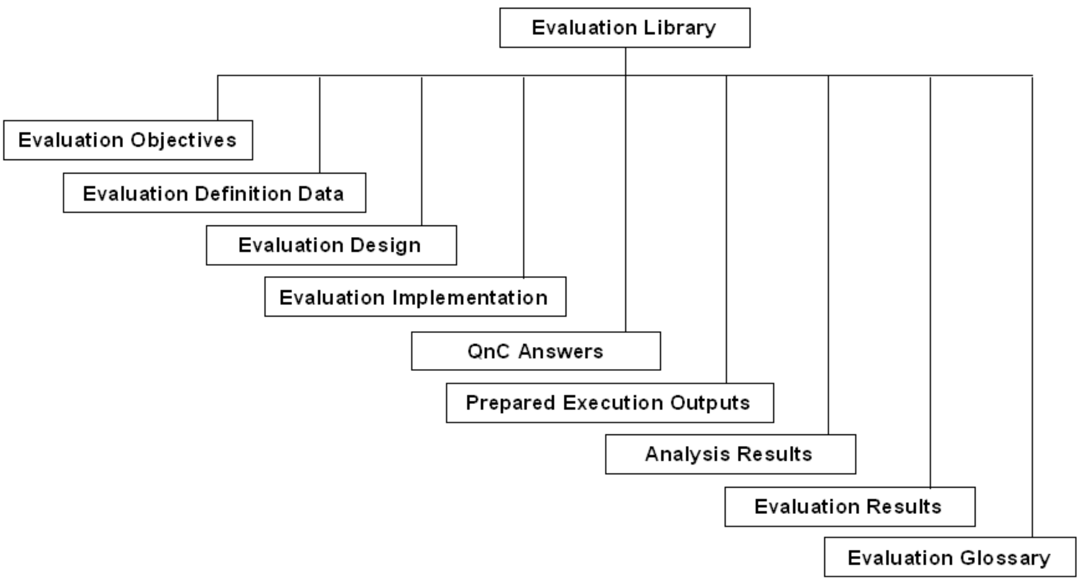


Figure 4: Evaluation Data Structure.

The aggregate SEDE data model has means for traceability of requirements and identification of dependencies to other SE information. This makes the dependency of for example the evaluation definition on a scenario item visible. Besides that it is possible to trace back from the evaluation results to the evaluation objectives by going from Evaluation Results to Analysis Results, from Analysis Results to Evaluation Definition and from Evaluation Definition to the Evaluation Objective. The Evaluation Data Structure is seen as an important asset in standardising the evaluation process. The standardisation of the used data format enables the reuse of evaluation data and encourages the Evaluation SMEs to spread their knowledge throughout the SE community.

The data structure is implemented in eXtensible Mark-up Language (XML) language that allows for the definition of actual data formats/data structures. By defining an open XML based data structure the independency and interoperability of the used tools is achieved. The evaluator or analyst can use his own familiar browser or XML editor to edit and access evaluation data. It is also possible to develop tailored tools for representing and processing the evaluating results as well as using the prototype evaluation tool set developed within the RTP11.13.

EVALUATION DEFINITION TOOL AND EVALUATION DEFINITION SELECTION TOOL

A lot of effort was used in getting to grips with the first activities of evaluation, i.e. the gathering of evaluation objectives and the definition of evaluation. To test the ideas and to give the evaluation SMEs the possibility to go into practise in these issues two prototype tools were developed: the Evaluation Definition Tool (EDT) and the Evaluation Definition Selection Tool (EDST). The EDT supports the evaluation SMEs and system engineers in collecting the user's objectives and defining the evaluation related aspects on basis of the objectives. The EDT comprises two editors:

- Evaluation Objective Editor for defining Evaluation Objectives
- Evaluation Knowledge Editor for deriving Evaluation Definitions

The EDT Framework is the basic component of the EDT. It provides a frame for and works as an interface between the EDT and external tools. The EDT has external interfaces for the following tools (see Figure 5):

- Evaluation Definition Selection Tool (EDST)
- Euclid RTP11.13 Repository

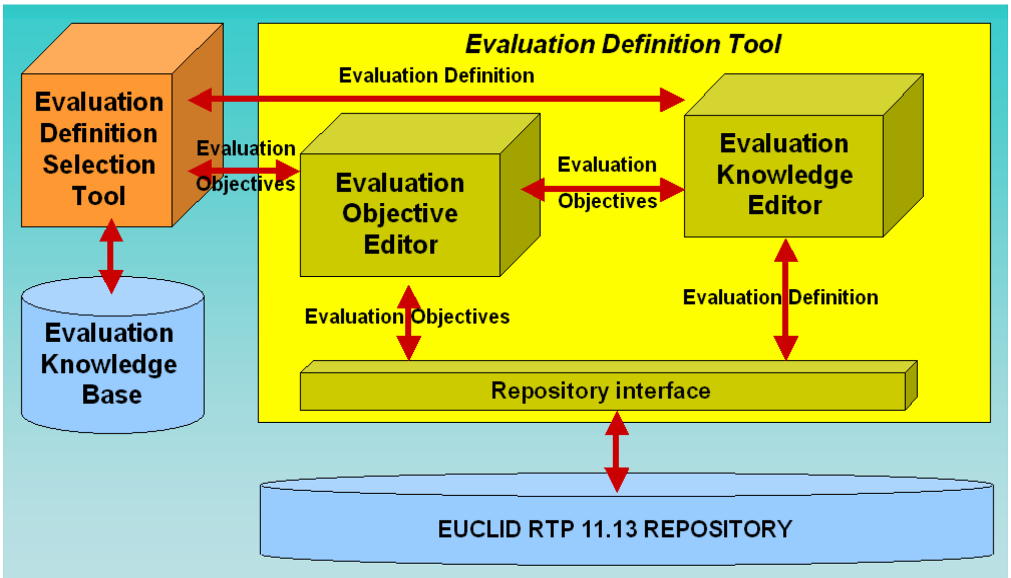


Figure 5: Evaluation Definition Tool Architecture.

The Objective Editor is a tool used for gathering the Evaluation Objectives. The objectives form a hierarchical tree-like structure. This structure allows to clearly present even a large number of objectives and to arrange the objectives by topic. The objective creation starts from the highest-level objective and is then elaborated in further detail until the objectives are specific enough to start the Evaluation Definition. The Evaluation Objectives are referenced to User Goals, which are an outcome of SEDEP step 0.

In the Evaluation Knowledge Editor (EK-Editor), the user is able to define the Criteria, Algorithms and Measures needed to assess the exercise according to the Evaluation Objectives. The EK-Editor also supports the use of questionnaires in the Evaluation. In the EK-Editor the user builds up so called Evaluation Definition Networks to represent the definitions. An Evaluation Definition Network consists of Inputs, Outputs and Assessment Nodes, which define the dependencies between Inputs and Outputs. The visualisation of the definition as a network eases the construction of the Evaluation Definitions by structuring the definition and enabling the reuse of network nodes (i.e. Algorithms and Measures). The Assessment Nodes represent the evaluation methods. Examples of evaluation methods are mathematical algorithms. The definitions of these may be linked to elements in the scenario, which are defined for the evaluated exercise. Criteria will be determined (according to the objectives) that are used to judge the quality of the collected data. In this case quality is related to the performance of the evaluated object and not e.g. to the data completeness. These evaluation criteria are also defined using the EK-editor.

The EDT also provides access to a pool of ready-made objectives and definitions through the EDST. The GUI of the EDST is presented in Figure 6. In the EDST it is possible to search for and select suitable evaluation objectives from a pool of ready-made objectives. On the basis of the selected objectives, the EDST searches related objectives. Artificial intelligence is used to search algorithms in an effective way. The objectives and definitions generated by the EDST are provided to the EDT, where they are attached to the existing data and can be edited if necessary. The EDST has its own knowledge base for utilising the searches. Naturally the EDST has facilities for inputting and editing the data in the knowledge base.

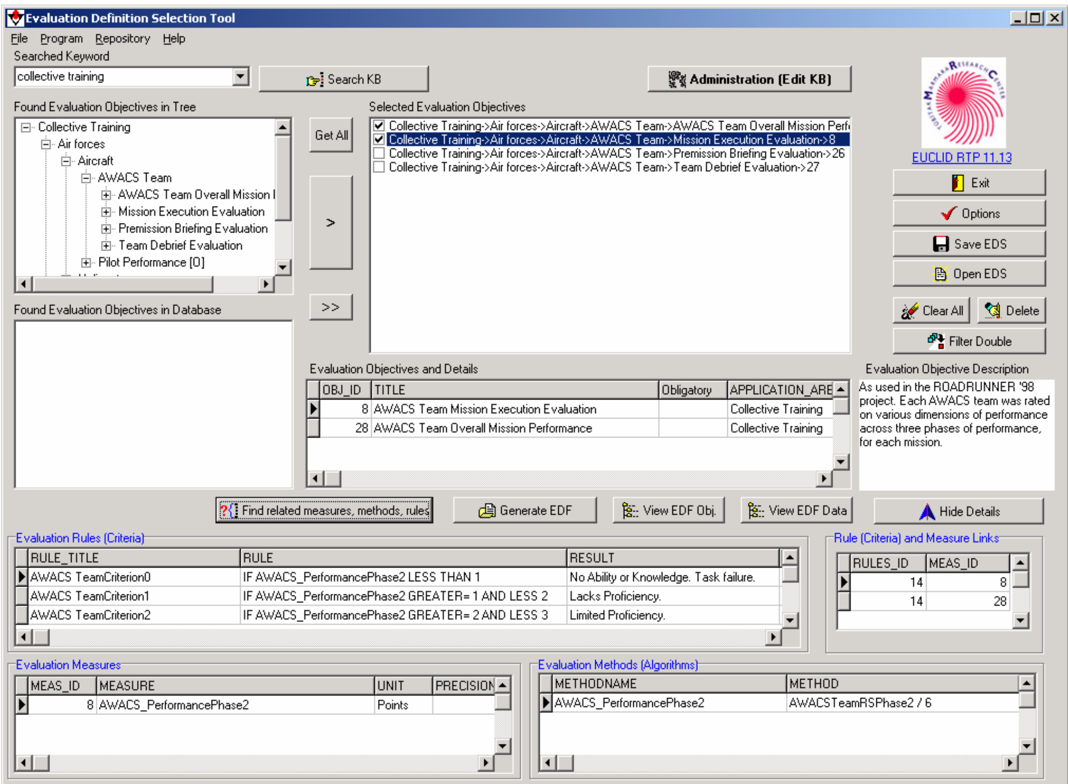


Figure 6: EDST.

The EDT may be used as an integrated part of the Synthetic Environment Development Environment (SEDE) or as a standalone tool. Using the EDT as part of the SEDE means, that the EDT

- Has access to the RTP 11.13 Repository
- Can be launched from the Synthetic Environment Management Tool (SEMT)
    - SEMT is an overarching tool developed for managing SE projects
- Stores data using the data format defined in the Evaluation Data Structure

When using the EDT as standalone tool the RTP11.13 Repository can not be accessed and thus it is not possible to link the evaluation information to neither user goals nor scenario. In the stand-alone mode the evaluation information is stored in the computer's local file system still using the Evaluation Data Structure.

EXECUTION EVALUATION TOOL

The Execution Evaluation Tool (EET) supports the user in post-processing the outputs acquired during the CT execution and in the analysis and evaluation of the results. This tool uses prepared execution outputs to apply the evaluation algorithms and criteria. The intention of the EET is to provide information on one hand needed to generate process feedback and corrective actions to improve the design and development of an SE and on the other hand needed to produce useful results for the evaluator to assess the trainees. The EET provides the user with structured analysis results for evaluation instead of lots of unstructured execution outputs. The EET provides the user with data in formats, which can be imported directly into documents or presentations.

The EET uses the evaluation definition (produced by the Evaluation Definition Tool) and the execution outputs from the Euclid RTP 11.13 Repository. It evaluates the CT exercise by processing the prepared execution outputs using the evaluation algorithms. The EET is in the same way as the EDT part of the SEDE. The EET is able to analyse the complete evaluation, fully automatic if the user wants to, but also gives the user the option to execute only a subset of evaluation algorithms, and, depending on the options per algorithm, the user may choose to adjust some of the algorithm's parameters. It may also be possible that a specific algorithm is available for interactive analysis. In that case the user will be able to interactively adjust (some of) the algorithm parameters and view the effect on the results immediately.

In Figure 7 the architecture of the EET is shown. A general evaluation may lead to the necessity for multiple commercial tools, because there is no single tool that can process all data. The tools 1, 2, and 3 in the figure may represent different tools from different vendors or the same tool (e.g. Matlab) used with different requirements on available tool boxes. To the user the EET will be presented as one tool, and the internal structure (e.g. the fact that there are multiple COTS tools at work) is hidden (the yellow box is surrounding all other components, including the COTS tools).

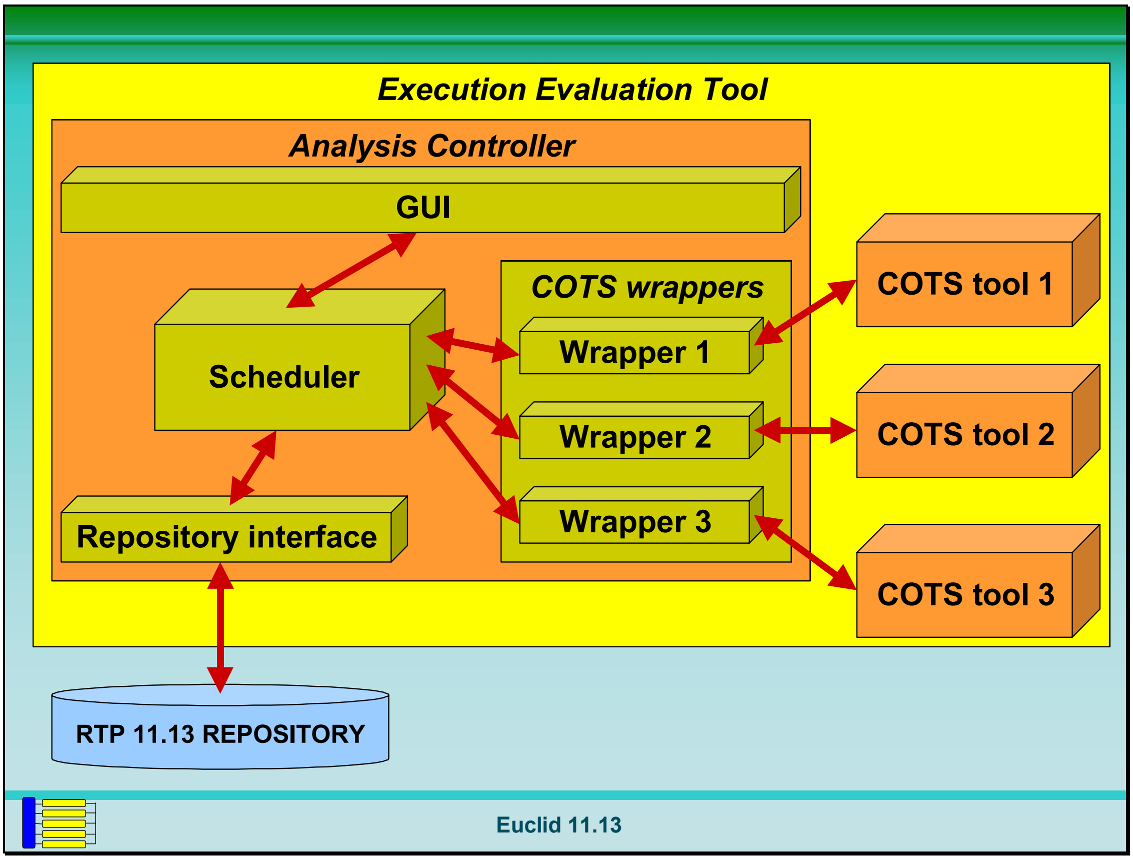


Figure 7: EET Architecture.

CONCLUSIONS

It is well known that in the training domain, individual skills are taught much more effectively using tailor-made training devices. The main advantage of SEs is in training certain collective skills. The distributed, group nature of Collective Training makes any measurement of performance difficult, with the level of difficulty increasing as the size of the unit under training increases. Currently only basic, usually subjective, forms of measurement exist at the sub-unit level and above. The EUCLID RTP11.13 team has identified this fact as a major obstacle for realising the potential of using Synthetic Environments (SEs) for CT purposes. The process of setting up SEs for CT needs to be refined further in order to derive more robust CT metrics, which could be used in live or synthetic exercises.

For this process the Federation Development and Execution Process (FEDEP) has been used as a baseline, but it has been found lacking on evaluation issues. One of the shortcomings was that the FEDEP does not cover the complete lifecycle of a SE. Therefore the FEDEP has been modified and extended with, among other things, a new step "Perform Evaluation". The resulting RTP 11.13 process is known as the Synthetic Environment Development & Exploitation Process (SEDEP). Discussions have taken place with the FEDEP Productisation Group, resulting to a total of 16 changes which are now included in the IEEE 1516.3 FEDEP version. The most important contribution of Euclid RTP 11.13 is the new FEDEP step 7 "Analyse Data and Evaluate Results".

The reuse of SE components, specifications, and definitions is a key factor to reduce the cost and time scale of creating and utilising SEs. To enable and facilitate the reuse of evaluation data Euclid RTP 11.13 has defined common data formats, captured in the SE Development Environment (SEDE) data model.

In this way, the data created by one tool can be read by the tool supporting the next activity in the process. All tools will have access to the RTP11.13 Repository, which will provide the mechanism for transferring data between them. The Data Interchange Formats (DIFs) defined by the HLA are being extended to other areas supported by the SEDEP. The data accessed by the SEDE can be from a local repository or from one that has been distributed over a Wide Area Network (WAN).

COTS tools are already available for supporting many of the SEDEP activities and three additional tools have been prototyped to support the evaluation related activities: the Evaluation Definition Tool (EDT), the Evaluation Definition Selection Tool (EDST) and the Execution Evaluation Tool (EET). The EDT supports the evaluation SMEs and system engineers in collecting the user's objectives and defining the evaluation related aspects on basis of the objectives. The EDT also provides via the EDST access to a pool of ready-made objectives and definitions. The EDST makes it possible to search for and select suitable evaluation objectives from a pool of ready-made objectives. The EET supports the user in post-processing the outputs acquired during the CT execution and in the analysis and evaluation of the results. The intention of the EET is to provide information needed on one hand to generate process feedback and corrective actions to improve the design and development of an SE and on the other hand needed to produce useful results for the evaluator to assess the trainees.

REFERENCES

[1] Training in a synthetic environment for improved operational effectiveness in collective air operations, Heather M. McIntyre & Ebb Smith, Defence Evaluation & Research Agency, I/ITSEC, November 2000, Orlando.

[2] Handbook of Simulator-based Training, Eric Farmer et al., Ashgate Publishing Ltd, 1999, Aldershot, UK.

[3] The Deliberate Application of Principles of Learning and Training Strategies within DMT, Winston Bennet Jr. and Peter Crane, NATO SAS-038 MTDS Symposium, April 2002, Brussels.

[4] High Level Architecture Federation Development and Execution Process (FEDEP) Model, Version 1.5, December 1999.

[5] The Euclid RTP 11.13 Synthetic Environment Development & Exploitation Process, K.J. Ford, Fall SIW, Paper 01F-SIW-124, September 2001, Orlando.

Validation Of Visual Threat Recognition And Avoidance Training Through Analogical Transfer

Dr. Barry P. Goettl
Air Force Research Laboratory
2485 Gillingham Drive
Brooks City-Base, TX 78235
USA

Alan Ashworth
Air Force Research Laboratory
2485 Gillingham Drive
Brooks City-Base, TX 78235
USA

Capt Ed McCormick (Ret.)
Air Force Research Laboratory
2485 Gillingham Drive
Brooks City-Base, TX 78235
USA

Michael Anthony
Galaxy Scientific Corporation
Blanco Road
San Antonio, TX 78232
USA

Summary

The Visual Threat Recognition and Avoidance Trainer (VTRAT) system addresses the need for realistic training that enables forces to respond quickly and operate effectively during threat engagements. This initiative combines state-of-the-art visual, interactive simulations with intelligent tutoring methodologies to train the visual scanners. VTRAT is an automated virtual intelligent instructional training aid, designed to introduce or refresh scanners on visual recognition of threats and on their duties during a threat engagement. This paper describes two studies that employ a backward transfer procedure for validating the training effectiveness of VTRAT. In experiment 1, 34 Air Force aircrew personnel were tested at Air Force Special Operations Command on their identification of surface-to-air missiles (SAMs). Several types of SAMs were simulated and presented to participants. Participants identified the model of SAM, and the accuracy of trajectory (whether or not the SAM would hit their simulated aircraft). Participants were assigned to either a novice group (no combat experience) or expert group (combat experience in either the Kosovo or Iraqi air campaigns). The experts outperformed the novices on all performance measures. In experiment 2, 39 aircraft scanners with the 19th Special Operations Squadron at Hurlbert Field, FL completed VTRAT training upon return from deployment in Afghanistan. There was a significant positive correlation between number of antiaircraft artillery (AAA) threats observed in theater and AAA threat recognition in VTRAT. Together these two studies provide evidence supporting the validity of VTRAT training for recognizing and avoiding visual threats. Additionally, survey results indicated that scanners rated the VTRAT visuals to be highly realistic and VTRAT training to be valuable compared to previous training methods.

Introduction

Air Force Special Operations Command (AFSOC) aircraft scanners (crewmembers) have the primary duty to identify antiaircraft threats, direct the pilot and crew in performance of evasive maneuvers, and deploy countermeasures during an antiaircraft threat engagement. Survival of slow-moving aircraft during enemy engagements with non-radar antiaircraft weapons hinges on the rapid detection and responses of visual scanners. Once enemies engage using optically guided antiaircraft artillery or infrared surface-to-air missiles (IR-SAM), scanners must detect the threat and respond with appropriate verbal calls dispensing countermeasures within a matter of seconds.

Traditionally aircrew members have received little or no training in detecting threats. Operating procedures are rarely practiced. It is impractical to train detection skills by firing live rounds or actual SAMs at aircraft during training flights. Moreover, while procedures for handling threat engagements are written down, they are not practiced often enough to become routine. In fact, many crewmembers may not be completely familiar with the specified procedures. Consequently, when facing threats in actual operations, crewmembers may not have the ability to respond in an accurate and timely manner.

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

In 1998 the Warfighter Training Research Division of the Air Force Research Laboratory (AFRL/HEAI), began development of a synthetic trainer for the 19th Special Operations Squadron (SOS). This training system, called the Visual Threat Recognition and Avoidance Trainer (VTRAT), was designed to train loadmasters in AC-130 Gunships to rapidly detect and execute countermeasures for various simulated threats (Figure 1). Basic system functionality, dictated by the prescribed training methodology, includes: (a) a view of the threat environment from the perspective of the scanner in the airframe, (b) accurate visualization of threat characteristics, (c) a high-fidelity interface for scanner communications and countermeasures, and d) dynamic feedback indicating timeliness, accuracy, and effectiveness of response.

Figure 1. Visual Threat Recognition and Avoidance Trainer (VTRAT).

The system combines state-of-the-art visual, interactive simulations with intelligent tutoring methodologies to provide an adaptive and efficient solution for training visual scanners. VTRAT decomposes the skill of threat detection and avoidance into subskills and focuses on each subskill. First the visual characteristics of various threats are displayed and discussed. Then learners are trained to recognize and distinguish enemy launches that pose an actual threat to the aircraft from those that do not. After learners achieve a level of proficiency on recognizing threats, the avoidance procedures are discussed and illustrated. Following this discussion, learners have an opportunity to practice these avoidance procedures. In this way the skill is built up from declarative knowledge into automated behavior.

Due to the success of the initial project, the VTRAT system was expanded to include all visual scanners on the AC-130 (Gunship), MC-130/U (Talon I), MC130/H (Talon II), MC130/P (Shadow), and the MH-53 (Pavelow). This expanded project required coding for each additional crew position viewpoint. In addition, each airframe was programmed to fly slightly different profiles and each was programmed for slightly different avoidance tactics. Moreover, the threat systems being modeled were certified on visual characteristics of antiaircraft weaponry, including missile fly-out, AAA rate-of-fire, trajectory, speed, etc., by the Air Force Information Warfare Center as valid for training purposes.

The goal of this research project is to validate the training effectiveness of the VTRAT system for training visual scanners to recognize threats and execute avoidance procedures. The most optimal way to test the validity is to measure the transfer from the training system to actual operations or actual exercise with live fire. Cost, risk, and logistics render this method of validation infeasible.

An alternative approach to validation is to measure the task transfer from experts to the training system. Knowledge and skill transfer is not unidirectional. Expertise gained during training transfers to the relevant task, but expertise gained on the relevant task will also transfer back to training. The critical assumption is that the fidelity of the training to the task is sufficient to promote transfer. Thus the representation of the necessary task knowledge must be sufficiently analogous to the task to promote transfer (e.g., see Singley & Anderson, 1989; Gick & Holyoak, 1980, 1983; Holyoak, 1984).

Measuring the extent to which skills and expertise acquired in field operations transfer back to training systems provides a training validation metric. This procedure is called backward transfer. The amount of backward transfer should reflect the fidelity of the training system (Wightman & Lintern, 1985). Goettl and Shute (1996) demonstrated how backward transfer could be used to improve the efficiency of part-task training. The backward transfer procedure is depicted in Figure 2. In their study Goettl and Shute trained one group using whol-task training and another group using a series of part tasks. Following training both groups performed the whole task and then completed the part tasks. We compared the performance of the whole-task group on the part tasks to the initial performance of the part-task group on the part tasks. Any part tasks that the whole-task group performed better (i.e., showed backward transfer) were identified as critical tasks and were included in a modified part-task condition. The modified part-task condition resulted in overall better training than whole-task training.

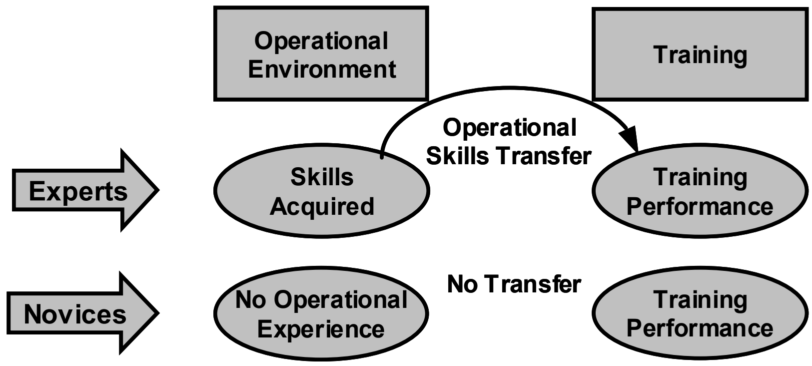


Figure 2. Backward Transfer approach to training validation.

In the present study backward transfer will be used to measure the training validity of the VTRAT system. The approach was similar to that employed by Goettl, Ashworth, and Chaiken (in press). Modular Control Equipment (MCE) operators conduct the same functions as Airborne Warning and Control System (AWACS) weapons directors except from ground-based stations. These duties and responsibilities include: control friendly fighter aircraft, monitor friendly fuel consumption, refuel friendly assets, monitor hostile aircraft, protect ground assets, deploy fighters against hostiles when necessary, collaborate with other operators, pass control of aircraft back and forth between sectors of the airspace. Goettl et al. had Air Force Modular Control Equipment (MCE) operators perform a synthetic team task configured to simulate the duties and responsibilities of a team of AWACS weapons directors. Operators performed the task under two different scenarios and with two different interfaces. Goettl et al. compared the operators' performance during acquisition and transfer phases with the performance of novices. They found that operators performed better during acquisition and transfer phases of the task. While novices showed dramatic performance decrements when the interface was changed, experienced operators showed no decrement at all. This clearly indicates that the operators had a richer knowledge of the AWACS task and were able to transfer that knowledge to the training task regardless of changes in the scenario and interface.

The present study presents two backward transfer experiments to validate the VTRAT system for training effectiveness. In Experiment 1 we compare the VTRAT performance of self-identified threat experts against the performance of threat novices. The prediction is that experts will perform better on VTRAT. In Experiment 2 the performance on VTRAT of aircrews will be correlated with the amount of threat experience during recent operations including Kosovo and Afghanistan. A positive correlation reflects positive transfer and will provide evidence of training efficacy. Aircraft scanners from the 19th SOS at Hurlbert Field, FL. completed VTRAT training upon return from deployment in Afghanistan. Correlations were calculated between experience with real-world threat engagements and VTRAT performance. In addition, we administered surveys to assess judgments of simulation realism and training effectiveness.

Experiment 1

Method

Subjects. Thirty-four Air Force Aircrew personnel were tested at AFSOC at Hurlbert Field, FL. Participants were assigned to either a Novice (n = 25) or Expert (n = 9) group based on their real-world military combat experience. The Novice group were Special Operations trainees, none of whom had combat experience. The Expert group had C-130 experience during either Operation Allied Force (Kosovo) or Operation Desert Storm (Iraq), and had engaged actual enemy air defenses.

Apparatus and Materials. The VTRAT is a state-of-the-art training system that incorporates high-fidelity graphics and voice recognition technologies (Figure 1). Five software components collaborate to provide VTRAT functionality: an Image Generator (IG), an Interaction Server (IS), a Threat System (TS), an Aircraft Host (AH), and a Situated Adaptive Learning Technology (SALT) component. These components were implemented using commercial-off-the-shelf (COTS) and government-off-the-shelf (GOTS) development tools on COTS hardware.

The SALT component allows both the automated instruction and an instructor to control the simulated environment. The instructor uses SALT's graphical user interface (GUI) to control the simulated threat environment and monitor student performance. The GUI allows the instructor to modify numerous simulation dynamics such as scanner viewpoint, aircraft altitude, airspeed. Currently, VTRAT contains courses to train threat recognition and avoidance for all duty positions requiring scanner training in the Gunship, Talon-I, Talon-II, Shadow, and MH-53J weapon systems.

The IG, TS, and AH form the heart of the 3-D visual simulation. These components run on an SGI Onyx2-IR2 and are certified High Level Architecture (HLA) compliant. The aircraft host models a C-130 or MH-53J, on a mission flight path, dynamically changing altitude, airspeed, and flight path as requested through scanner communications, the instructor's GUI, or automated instruction. The TS models a variety of AAA and SAMs. It dynamically generates threats based on instructional objectives that dictate threat placement and targeting parameters.

The student views the appropriate crew position perspective of the simulated threat environment on a high-resolution 67" display system. The student wears a headset, with a microphone attached for purposes of voice recognition, and holds a communication control and a flares countermeasures control. Both controls are from the actual aircraft, modified to communicate with VTRAT through a PC game port. The IS (running Windows 2000 on a 500 MHz PC) delivers instructional text as synthesized speech output, and monitors scanner communications. The speech recognition software provides speaker-independent, continuous speech recognition, with no training requirements.

Procedure. In the first phase trainees were required to discriminate between the three SAMs until they reached an accuracy criterion. They also had to state whether or not the SAM was accurate or inaccurate (would or would not hit the C-130). There was no time pressure. Both the identity and accuracy of the SAMs were evenly distributed across trials. To successfully complete this phase trainees had to complete at least 20 trials, and 9 of the last 10 trials had to be correct for both identity and accuracy of SAM. Trainee's verbal responses were processed by the speech recognition system, which then supplied accuracy feedback. The second phase was identical to the first except that it was conducted under time pressure. The trainees had to meet criterion first within an 8-second time window, then a 6-second window, and finally within a 4-second window. Thus not responding within the time window became an additional type of error.

Results

Reported here are the data from the final phase of training. Independent samples t-tests revealed four main effects (Table 1). Due to unequal group size, homogeneity of variance was not assumed. There was a significant difference in Proportion Correct, $\underline{t}$ (12.58) = 2.49, $\underline{p} < .05$, indicating that Experts ($\underline{M}$ = .78) had a higher proportion correct than Novices ($\underline{M}$ = .68). There was a significant difference in total number of SAMs shown, $\underline{t}$ (26.69) = -2.22, $\underline{p} < .05$, indicating that Novices (M = 32.28) required more trials to reach criterion than Experts (M = 24.33). There was a significant difference in the number of Time Outs, t (17.02) = -2.05, p = .05, indicating that Novices (M = 5.44) were more likely to miss the time window for responses than were Experts (M = 2.89). Finally, there was a significant difference in the overall minutes to finish training, t (31.09) = -2.00, p = .05, indicating that Novices took longer (M = 16.56) than Experts (M = 12.59).

Dependent Variable	Novice	Expert	Significance
Proportion Correct	.68	.78	< .05
Total Number Shown	32.28	24.33	< .05
Time Outs	5.44	2.89	< .05
Total Minutes to Criterion	16.56	12.59	< .05

Table 1. Significant VTRAT outcome measures with means.

Discussion

The superior performance of the Expert group suggests that experts are in fact transferring their extant knowledge from their real-world SAM experience to the VTRAT training task. It can then be inferred that there is sufficient fidelity between VTRAT and its real-world counterpart to make VTRAT a viable training device.

It is often the case that training systems cannot be validated through their real-world counterparts. However, this does not preclude other methods of validation. Measuring the transfer from experts' extant knowledge back to the training system is an affordable and theoretically appropriate method of validation in such circumstances.

Considering the high costs in human life, materiel, and mission effectiveness associated with ineffective training systems, there is ample reason why Air Force training systems such as VTRAT should always be validated.

Experiment 2

Method

Participants. Participants were 39 Aircrew members from the 19th and 20th Air Force SOSs. Participants ranged in age from 18 to 30 and had a high school diploma or Graduate Equivalency Degree (GED). Participants included pilots, co-pilots, flight engineers, loadmasters, scanners, and gunners from five different airframes: gunship, Talon I, Talon II, shadow, and MH-53. All participants were recruited by AFSOC personnel and participated voluntarily.

Equipment. The equipment used in Experiment 2 is the same as was used in Experiment 1.

VTRAT Task. In this Experiment participants completed two drill lessons of VTRAT. In one drill participants had to detect and initiate avoidance tactics for different IR SAM threats. The threats modeled in the system were three MANPAD systems including the SA-7, SA-14, and SA-16. Participants had to make the appropriate call and deploy countermeasures within a specified length of time. The drill consisted of 20 trials. In the second drill, participants had to detect and initiate avoidance tactics for AAA threats. Participants were graded on each component of the avoidance call and whether they completed the call within a specified window. As part of the detection, participants had to determine whether the AAA threats were firing at the aircraft or not, and whether the maneuver they called was appropriate. Participants completed a minimum of 20 trials and had to get 9 of the last 10 to complete the drill. Five different AAA threats were modeled: 14.5 mm, 23-2mm, 23-4mm, 37mm, and 57mm.

Questionnaire. The questionnaire consisted of three parts. The first part of the survey assessed general mission experience as well as combat experience. Participants indicated number of flights and total flight hours in theatre. In addition they were asked to indicate how many times they had observed SAM and AAA threats firing at them or other aircraft in theater.

The second part of the survey dealt with specific threats modeled in VTRAT. Participants were asked to indicate how often they had observed each of the different threats modeled in VTRAT and to rate the VTRAT simulations on different dimensions. Rating scales consisted of a 6-point Likert–type scale with anchors representing very inaccurate (1) and very accurate (6). Only those having seen the specific threats in theater were required to rate the simulations.

In the final section of the survey participants were asked to rate the quality of VTRAT training compared to previous training methods. Participants rated the quality of the discussion of the visuals, discussion of the tactics, and realism of the threats and scenarios. Ratings were made using a 6-point Likert-type scale with poor (1) and excellent (6) as the anchors. The aspects of VTRAT that were rated included (a) discussion of visual features of AAA threats, (b) discussion of visual features of IR-SAM threats, (c) discussion of avoidance tactics for AAA threats, (d) discussion of avoidance tactics for IR-SAM threats, (e) realism of AAA threats, and (f) realism of IR-SAM threats.

Procedure. The data reported here were collected at the 19th SOS in Hurlbert Field, FL. Participants signed in and were given a subject-matter expert (SME) number to maintain confidentiality. They were briefed on the goals and purposes of the study. Following these instructions, participants completed the IR-SAM recognition and avoidance drill. Next, participants completed the AAA threat avoidance drill. Finally, participants were asked to complete the questionnaire, were debriefed, and released.

Results

Of the 39 participants only 33 completed the drills. Due to computer errors, data from 3 participants were lost. Of the remaining 30 participants with performance data, there were 16 gunship crewmembers, 3 Talon crewmembers, 6 Shadow crewmembers, and 5 MH-53 crewmembers.

Performance and Experience. The primary goal of this study was to validate the accuracy of the simulation. The approach was to correlate accuracy of threat calls with amount of experience. If the visual characteristics of the threat simulations and flight simulation are accurate, there should be a significant positive correlation between performance on the threat calls and amount of experience. In the present study, experience was measured multiple ways: number of missions in theater, number of hours in theater, and number of times AAA or SAMs were seen fired upon self or others. On average, participants viewed AAA shot at themselves 3.67 time (SD = 6.12), AAA fired at others 3.07 times (SD = 3.859), SAMs shot at themselves 0.33 times (SD = 0.84), and SAMs fired at others 0.23 times (SD = 0.73).

Each of these measures was correlated with performance on threat. Table 2 shows the correlations between experience measures and accuracy of AAA and SAM threat calls and locations.

Experience	**AAA Threat**	**AAA Location**	**SAM Threat**	**SAM Location**
Total Missions	.155	.036	-.128	-.132
Total Hours	.057	-.276	.133	-.053
Threat at own ship	.116	.227	.109	-.042
Threat at other ships	.447*	.301	.184	-.102

Table 2. Correlations between VTRAT performance and threat experience: number of missions, number of mission hours, number of times viewed threat fired at own aircraft, and number of times viewed AAA fired at other aircraft (*p < .05).

As indicated in Table 2, in general, performance on AAA threat calls in VTRAT are positively correlated with experience. Moreover, correlations tend to be stronger for the experience metrics based on frequency of viewing actual threats in theater. In spite of this positive evidence, only the correlation between threat call and number of times viewed AAA fired at others reached statistical significance.

Correlations involving SAM performance and experience, as indicated in Table 2 were not as strong as correlations for AAA. This is likely due to a restriction of range problem. That is, very few participants had viewed actual SAM threats in theater. Additionally, performance on SAM threat calls were highly accurate ranging from 95% correct for shadow crews to 98 % correct for gunship crews. Nevertheless the correlations between threat call accuracy and the number of times SAMs had been viewed are positive.

Subjective Ratings of Training Value. An important metric in training is the perceived value of training. Participants were asked to compare the VTRAT training to previous training approaches on several dimensions. These dimensions included discussion of visuals and tactics for AAA and SAM threats, and realism of AAA and SAM threats. Table 3 shows the average ratings for several VTRAT dimensions.

VTRAT Dimension	Average Rating
Discussion of AAA Visuals	5.48
Discussion of SAM Visuals	5.46
Discussion of AAA Tactics	5.04
Discussion of SAM Tactics	5.35
Realism of AAA Threats	5.27
Realism of SAM Threats	5.24

Table 3. Average ratings of VTRAT quality metrics

As indicated in Table 3, overall the ratings of VTRAT were relatively high. On a 6-point scale where 6 is excellent, no rated aspect of VTRAT was rated below a 5.00. Compared to other training approaches that participants experienced, VTRAT was rated nearly excellent on all dimensions. The discussion of visuals showed the highest ratings followed by the discussion of SAM tactics, then realism of the threats. The aspect of VTRAT rated the lowest was the discussion of AAA tactics. This is not entirely surprising given that avoidance of AAA threats is more complex and based on perceptual judgments that are difficult to model on a 2-dimensional display.

Discussion

The data presented here support the conclusion that VTRAT provides valuable training experience to visual scanners. This evidence comes not only from the subjective ratings of visual scanners but from backward transfer analysis.

Perhaps the most important finding in this study is the strong positive correlation between experience and VTRAT performance. Accuracy of AAA threat calls was positively and significantly correlated with the number of times scanners had viewed AAA threats firing at other aircraft. This is the kind of backward transfer evidence that supports the validity of training. Those with more experience and familiarity with actual threats performed better on recognition and avoidance drills. While in the current data, the strongest effects were shown for AAA performance, correlations between SAM threat data and experience were positive for all experience measures except total missions.

The difference in backward transfer between AAA experience and SAM experience may reflect the relatively little experience participants had with SAM threats. Moreover, the experience with SAM engagements was not correlated to total number of missions (.116 and -.050 for SAMs at you and at others, respectively). For AAA experience the number of total missions was strongly correlated with number AAA threats viewed (.626 and .517 for AAA fired at you and at others, respectively). Again the weak correlations likely reflect restriction of range problems. Very few participants had seen SAMs at the time of testing (M = .33).

Finally, we obtained strong positive results from the subjective ratings of the training value. All rated aspects of VTRAT training received an average rating of 5 or more on a scale where 6 represented excellent value. These data reflect the general response that VTRAT developers have received from AFSOC personnel. In fact, the system was so highly valued that AFSOC personnel created compact disks of VTRAT simulations and passed them around to aircrews during the Afghanistan operations. In summary, the data presented demonstrate and confirm the validity of VTRAT simulations and drills for training the recognition and avoidance of visual threats.

General Discussion

These two experiments provide converging evidence that the VTRAT system is effective for training the recognition and avoidance of threats. Both experiments used a backward transfer approach to training validation. This technique measures the amount of transfer from the real-world task to the training system. This approach assumes that if the training system captures important task dimensions, then experts will perform better on the training system than will novices. In Experiment 1 backward transfer was measured by comparing VTRAT performance of operators with combat experience to performance of operators without combat experience. In that experiment experts performed better on all performance metrics of IR-SAM detection in the VTRAT. This finding supports the conclusion that experts were able to transfer combat experience to the training environment.

In Experiment 2, experience was measured more directly through a survey measuring specific experience perceiving AAA and SAM threats. In this experiment, the extent of real-world experience viewing AAA was positively correlated with VTRAT performance in detecting AAA threats. The finding that operators with greater experience viewing actual AAA threats performed better than those with less experience supports the conclusion that VTRAT captures important dimensions of the threat detection task.

It is difficult to validate the effectiveness of simulators designed for combat training. The most logical approach for measuring transfer from the training to the operational task is not practical. Measuring transfer from the training system to a very high-fidelity simulation can be expensive and is limited to the validity of the simulation. The backward transfer approach provides a practical inexpensive way of validating the effectiveness of military training systems. The present paper demonstrates two approaches to backward transfer to the validation of a training system and can serve as a model for future validation efforts.

Acknowledgements

The Air Force Research Laboratory gratefully acknowledges Command Technologies Incorporated for providing the technical implementation for VTRAT. Carol Horwitz (Team Lead), Keith Brown (Technical Lead), and programmers Doug Estrumse, Steve Holley, Esteban Garza, and Neil Kalinowski were primary forces in making VTRAT a state-of-the-art visual trainer. We also acknowledge the support of LtCol Bill Wimpee AFRL/HEAI; Capt Ed McCormick, AFRL/HEAI; and MSgt Ken Taylor, AFSOC.

References

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. Cognitive Psychology, 12, 306-355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. Cognitive Psychology, 15, 1-38.

Goettl, B.P., Ashworth, A.R.S., III, & Chaiken, S.R. (in press). Advanced distributed learning for team training in command and control applications. In Fiore and Salas (Eds.), *Where is the learning in advanced distributed learning?* Washington, DC: American Psychological Association.

Goettl, B.P., & Shute, V.J. (1996). An analysis of part-task training using the backward-transfer technique. Journal of Experimental Psychology: Applied, 2, 227-249.

Holyoak, K. J. (1984). Analogical thinking and human intelligence. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence, (Vol 2, pp. 199-230). Hillsdale, NJ: Erlbaum.

Singley, M. K., & Anderson, J. R. (Eds.) (1989). The transfer of cognitive skill. Cambridge, MA: Harvard Press.

Wightman, D.C., & Lintern, G. (1985). Part-task training for tracking and manual control. Human Factors, 27, 267-283.

Airmanship Training For Modern Aircrew

Ms. Louise Ebbage
BAE SYSTEMS Advanced Technology Centre
Sowerby Building, FPC 267, PO BOX 5,
Bristol BS34 7QW,
United Kingdom

Mr. Phil D. Spencer
BAE SYSTEMS Training Solutions
Warton Aerodrome, W295E, Mill Lane,
Preston, PR4 1AX
United Kingdom

SUMMARY

It is widely accepted within aviation circles that *airmanship* is key to modern aviation, yet there is considerable confusion as to what airmanship actually comprises. To some it is the "stick and rudder" skills associated with manual flying; to others it is the cognitive skills associated with decision-making and judgement. There is also a common belief that airmanship cannot formally be trained and is simply acquired through experience. None of these are completely true on their own. Instead, airmanship is a mixture of all of these attributes and much more besides.

This paper explores the concept of airmanship — its definition and basic components, and explains the importance of taking a holistic view of airmanship training as the necessary foundation for the creation of competent and professional aviators. It also presents a series of practical guidelines that can be applied in the training and assessment of airmanship skills.

Keywords: airmanship, training, cognitive skills, discipline

INTRODUCTION

The advent of modern technology in the cockpit has created a shift in the skills demanded of military aircrew, testing their ability to manage complex systems and to cope with unprecedented levels of information and data. Cognitive skills such as decision-making and situational awareness now take on an almost overriding significance and are fundamental to achieving a battle-winning edge. A step change in the conduct and content of military training is required to meet these new operational demands.

To meet this challenge, the UK Ministry of Defence has begun a 10 year programme to modernise military flying training across the UK's Armed Services. The UK Military Flying Training System (UKMFTS) will provide a modern, holistic approach to aircrew training that will meet the future demands of the UK front line. As part of the changes wrought by the UKMFTS it is expected that considerable emphasis will be placed upon the employment of the latest teaching technologies and training platforms. However, perhaps the greatest change will be the increased emphasis that will be placed on developing *airmanship*, because it has been recognised to be the key element in producing outstanding aircrew performance.

Although the importance of airmanship has long been undisputed, there remains considerable confusion as to what it actually is and how it is best taught. This paper examines the inherent qualities associated with effective airmanship, identifies the key knowledge, skill and attitude requirements and proposes a number of strategies for the training and assessment of airmanship skills.

DEFINITIONS OF AIRMANSHIP

Airmanship is a term that has relevance throughout aviation from commercial and general aviation through to the military domain. It applies equally to pilots and non-pilot aircrews (e.g. navigators) and is as relevant on the ground as it is in the air. Airmanship is accepted as being extremely important, yet it is a concept

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

that has been difficult to define — it is something that all aircrew understand but seem unable to put into words.

To some, airmanship simply means developing expert flying skills; to others it means exercising good judgement. To some, it is a collective term used to summarise all the skills and knowledge required to operate an aircraft; to others is simply represents a professional attitude or code of conduct.

Many researchers have attempted to define airmanship but as yet there is no universally accepted definition. Those offered in the literature include:

"Airmanship is effective decision making to support a sequence of actions."	Training Development Support Unit 2000 [1]
"Airmanship is the care and attitude that you bring to the conduct of your flying. It encompasses consideration for your passengers, care of your aircraft, courtesy to other airspace and airfield users and the self discipline to prepare and conduct your flights in the most professional manner possible. It is not just flying skill that distinguishes a good pilot; it is his or her standard of airmanship."	The Aviation Theory Centre 2001
"Airmanship is a personal and situational management state required to allow a human being to enter and exit, in safety, an environment which they were not naturally designed to inhabit."	Hayes 2002 [2]
"Airmanship is the consistent use of good judgement and well developed skills to accomplish flight objectives. This consistency is founded on a cornerstone of uncompromising flight discipline and developed through systematic skill acquisition and proficiency. A high state of situational awareness completes the airmanship picture and is obtained through knowledge of one's self, aircraft, team, environment, and risk."	Kern 1996 [3]

From the literature, it is difficult to establish whether airmanship is a process, a state, a skill or an outcome. Whether it means having expert flying skills, sound judgement or good situational awareness; or whether it simply means having the "right attitude".

The confusion stems from the fact that airmanship is *all of these things*. Airmanship is a multi-dimensional concept that involves acquiring and exercising both cognitive and physical skills in consonance. Moreover, it is about having the self-awareness and motivation to improve skills that may be lacking, and is a personal state or mind-set that compels aircrew to approach each flight with discipline and an appropriate attitude balancing safety against mission objectives.

Airmanship is also about achieving a balance. An airman who exercises good judgement but lacks the dexterity of control needed to operate an aircraft safely will not excel in airmanship. Similarly, an expert flyer (or operator) will fall short if he adopts a risk-taking attitude.

In essence, airmanship is about exercising judgement, discipline and having expert control of an aircraft and a situation.

Hence, airmanship can be defined as follows:

> A personal state that enables aircrew to exercise sound judgement, display uncompromising flight discipline and demonstrate skilful control of an aircraft and a situation. It is maintained by continuous self-improvement and a desire to perform optimally at all times.

THE AIRMANSHIP APPROACH

Airmanship is more than simply having the requisite knowledge and skills; it is also about having an appropriate attitude, self-discipline and a desire to perform optimally at all times. Airmanship is an *approach* to aviation, which manifests itself in excellent performance.

The essence of the airmanship approach is captured in the model shown in Figure 1.

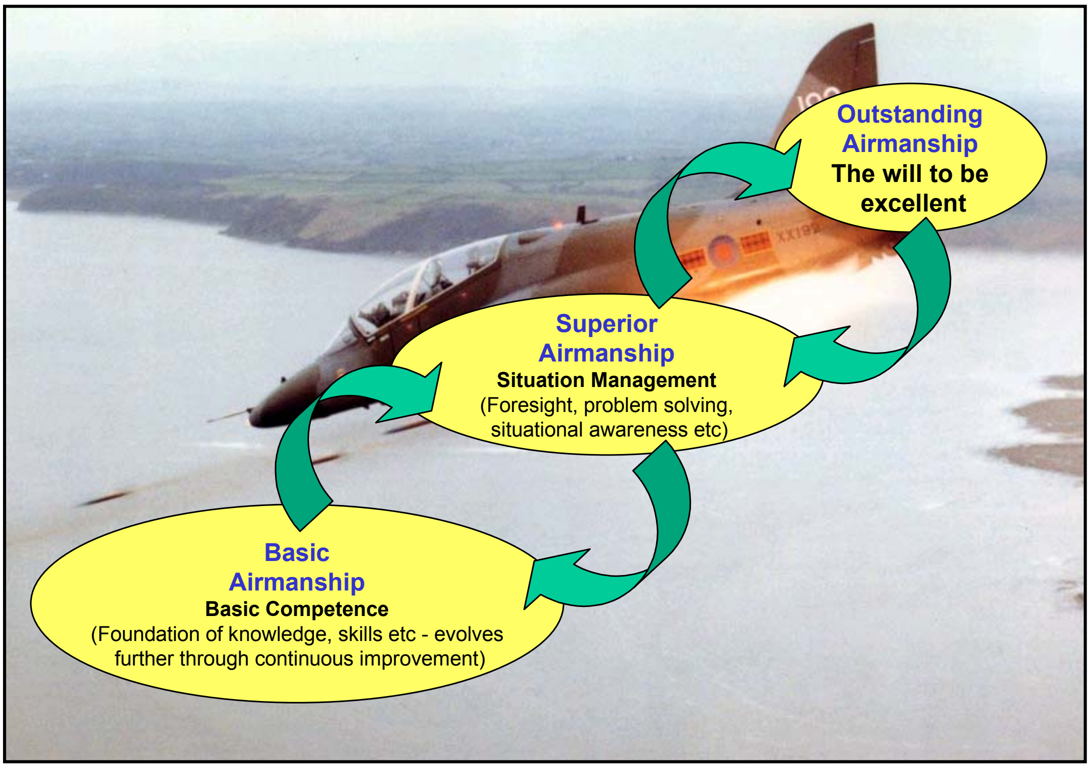


Figure 1: Levels of airmanship [4]

Aircrew operating at the basic level have the necessary knowledge, skills and attitudes and exhibit a textbook-based performance. With additional motivation, knowledge and experience, aircrew can move up to the superior level. Here, aircrew do more than simply follow standard operating procedures — they use foresight to anticipate problems and use higher-order skills such as situation assessment, judgement and problem solving to take a proactive rather than reactive approach to situation management. The demands of modern aviation necessitate that aircrew attain a superior level of airmanship.

At the highest level, is the desire to achieve excellence in all aspects of performance. Those operating at this level are dedicated to self-improvement and have a genuine desire to perform optimally at all times. Aircrew operating at this level seek airmanship excellence and this manifests itself in outstanding performance.

INDICATORS OF EFFECTIVE AIRMANSHIP

There are many examples of *ineffective* airmanship in the literature, but there are somewhat fewer examples of *effective* airmanship. However, those examples that can be found are extremely useful in making explicit the qualities associated with superior airmanship.

Take for example, the crew of a Delta Airlines flight from Houston to Dallas who demonstrated superior airmanship by landing the aircraft safely despite smashing into a flock of birds that destroyed one engine, damaged another and caused serious airframe damage.

> An air traffic controller had asked the crew to participate in a 'no airspeed restriction' test being run by the FAA. The 727 accelerated, as requested, and at 6,000 feet struck a flock of snow geese. The crew instantly found the aircraft vibrating intensely and all power was lost in one of its three engines. The first officer's cockpit instruments had also failed, and the noise in the cockpit was deafening. The crew worked as a team to return the crippled aircraft to Houston. The first officer flew using the captain's instruments, while the captain, second officer and line check second officer analyzed the situation and performed the appropriate emergency procedures. They declared an emergency in the air with ATC and informed passengers of their situation. With the captain taking the controls on the aircraft's final approach, they landed safely with no injuries. [6]

The overriding theme running through examples of effective airmanship is the ability of aircrew to "control" a situation by using both their training and a certain amount of on-the-spot ingenuity. Specific qualities associated with effective airmanship include the following [4]:

- **Discipline** - abiding by procedures, despite the peculiarity of the situation.
- **Communication** - keeping others (e.g. ATC) informed of developments.
- **Teamwork** - working well together to resolve problems and maintain control.
- **Knowledge** - having a deep understanding of aircraft systems and operation.
- **Expertise** - transfer/retention of knowledge and skills.
- **Situation Assessment** - analysing and assessing unusual developments.
- **Judgement** - calling upon prior training and expertise to resolve unusual problems.
- **Decision Taking** - taking decisive action.
- **Resource Management** - allocating resources to ensure control of the larger situation is maintained whilst specific problems are being addressed.
- **Goal Prioritisation** - prioritising safety above personal concerns.

These attributes of airmanship emerge once something has gone badly wrong, however, good airmanship also means preventing things from going wrong in the first place, and so to the above list we can add:-

- **Situational Awareness** - maintaining awareness; being alert to any unforeseen situations arising.
- **Foresight** - anticipating potential hazards.
- **Planning** - working out courses of action to deal with potential hazards.

The above attributes suggest that there are explicit knowledge and skills that expert airmen employ to bring about positive outcomes to adverse situations. However, many of these skills and knowledge already form the basis of aircrew training programmes, particularly in the civil sector. So why is it then that some aircrew demonstrate superior airmanship when others do not?

The main discriminator is an airman's attitude towards aviation.

The Airman is a person who maintains a valid skill and knowledge currency such that when the unexpected does happen there is ability and composure enough to manage the situation into safety. He or she is a person with a sense of balance and intelligence enough to heed the lessons of the past, apply them in the present, and so ensure a future to be able to fly again, and again, and again.[2]

THE FOUNDATIONS OF AIRMANSHIP

Like intelligence, airmanship is often regarded as an innate quality — a natural ability that some aircrew posses and others do not. This perspective makes it extremely difficult to develop a working definition of the concept and fosters the belief that airmanship cannot be taught.

Whilst it is true that airmanship is a personal quality and individuals do vary in their natural ability for it, the assumption that the concept cannot be taught is false. Much is known about the underpinning elements that define airmanship and with training and encouragement pilots can be motivated to seek airmanship excellence.

Various studies have attempted to determine the essential ingredients of airmanship; perhaps the most comprehensive to date is a study conducted by Kern (1996) in which a model was developed that described the main ingredients of airmanship.

A more simplistic model is presented in Figure 2 that defines more specifically the foundations of airmanship.

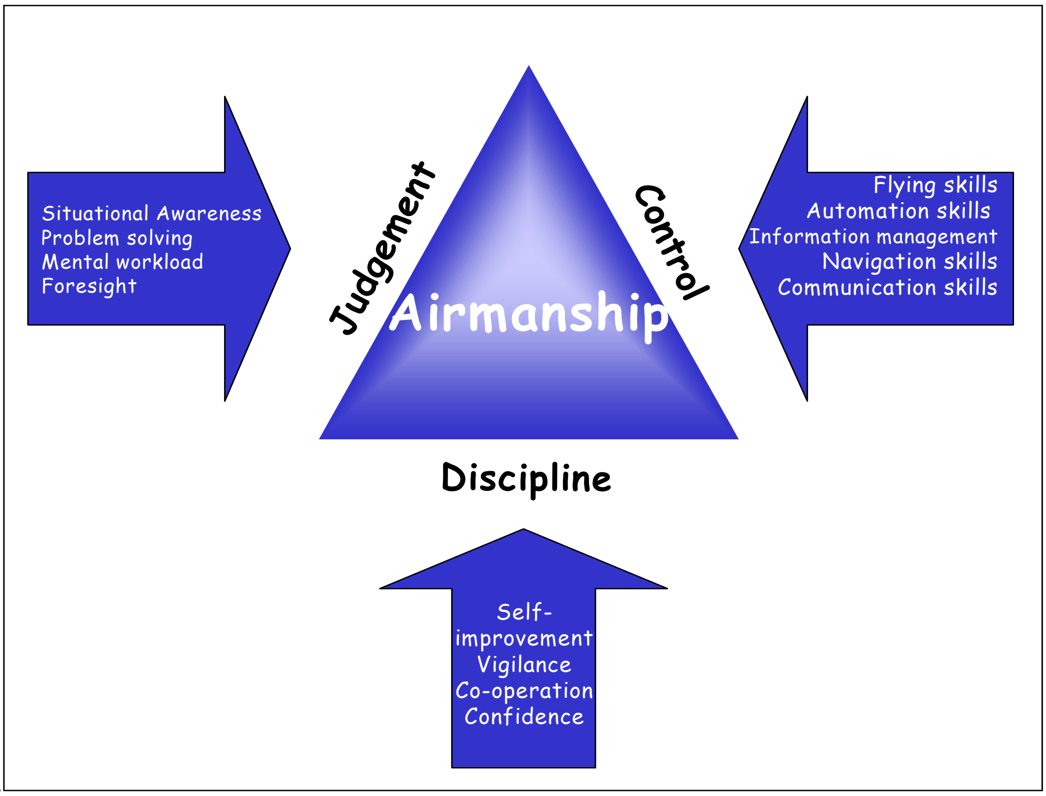


Figure 2 The elements of airmanship [5]

Judgement is used in the broadest sense to emphasise the need for aircrew to make conscious, intuitive, timely and well-founded decisions. *Control* is used as a reminder to maintain control of an aircraft whilst evaluating a situation and to execute a planned course of action with precision and accuracy — many good judgements are undone by failures in execution. *Discipline* is needed to detect potential errors at the earliest opportunity and to formulate considered judgements and execute controlled actions.

The foundations of airmanship are built on a specific set of knowledge, skills and attitudes (see Table 1). The elements listed in Table 1 should form the core syllabus of a training programme teaching the *foundations* of airmanship.

<table>
<tr><th colspan="3">THE FOUNDATIONS OF AIRMANSHIP</th></tr>
<tr><th>KNOWLEDGE</th><th>SKILLS</th><th>ATTITUDES</th></tr>
<tr><td>Knowledge of aircraft
- Deep understanding of aircraft sub-systems, emergency procedures, cockpit automation, aircraft flight characteristics and operating limits.</td><td>Physical skills
- Flying skills
- Navigation skills
- Instrument flying
- Emergency handling / recovery
- Combat survival</td><td>Hazardous attitudes
- Understanding the five main hazardous attitudes, the antidotes and the impact on airmanship (see Table 2)</td></tr>
<tr><td rowspan="2">Knowledge of environment
- Understanding the physical environment and the effects on aircraft control.
- Understanding the regulatory environment.
- Understanding the organisational environment and the challenges posed to airmanship.</td><td>Cockpit management skills
- Avoiding the pitfalls of automation (over-reliance, complacency, bias)
- Information management skills</td><td>Professionalism
- Understanding the values and principles embodied in airmanship.</td></tr>
<tr><td>Communication Skills
- Vigilance in monitoring communications
- Using appropriate communication (phraseology, clear, concise)
- Active listening
- Inquiry through communications</td><td>Self-improvement
- Developing the motivation needed for life-long learning
- Understanding the requirement for self-assessment in flight.
- Developing the will to achieve performance excellence</td></tr>
<tr><td rowspan="2">Knowledge of risk
- Understanding the risks to discipline, skill and proficiency, knowledge, SA, judgement, aircraft, self.</td><td>Cognitive skills
- Understanding and maintaining situational awareness
- Problem solving / decision-making skills
- Understanding and managing workload
- Self-assessment</td><td rowspan="2">Discipline
Discipline in terms of:
- flight preparation
- flight discipline (e.g. vigilance/ look-out, SA maintenance, operational & regulatory policy)
- knowledge & skills maintenance
- post-flight evaluation
- self-discipline (managing stress, managing attitudes)</td></tr>
<tr><td>Team skills
- Performance monitoring
- Leadership / initiative
- Interpersonal skills
- Co-ordination & decision-making
- Team communication and SA</td></tr>
</table>

Table 1: The knowledge, skills and attitudes that are the foundations of airmanship [5]

However, to achieve outstanding airmanship, aircrew must also adopt the values and principles that embody the airmanship philosophy. Hence training also has a role in shaping an airman's entire approach to aviation.

TRAINING AND EVALUATING AIRMANSHIP

Airmanship has traditionally been acquired as a by-product of conventional training. It is developed by exposing trainees to plausible scenarios and discussing key airmanship points. The increasing number of accidents attributed to human error led to the introduction of dedicated training programmes to improve pilot judgement and co-ordination. However, airmanship relies on the integration of physical and cognitive skills, with knowledge and discipline and should therefore be trained holistically. This means that soft skills training should be integrated within "conventional" training rather than singled out for specialist treatment.

Training airmanship requires three essential elements. First, instructors must explain the basic concept about why airmanship is so important and the rewards it offers. This is necessary to embed the values and motivation needed to achieve airmanship excellence. Secondly, instructors must teach the knowledge, skills and attitudes that are the foundations of airmanship. Aircrew should also be provided with sufficient performance data, learning material and training facilities to enable them to continue learning between formal training sessions, thus fostering a culture of continuous learning. Finally, airmanship must be assessed thoroughly and objectively to provide feedback on the knowledge and skills that must be improved.

The following guidelines have been developed to assist in the design of airmanship training material and training programmes. The guidelines are divided into four main categories:

1. Introducing the concept
2. Overall teaching strategies
3. Teaching the foundations of airmanship
4. Assessing airmanship

INTRODUCING THE CONCEPT

1. Provide real-world case studies to demonstrate the importance and rewards of airmanship.
2. Use models of airmanship to present the foundations of airmanship and the levels of airmanship competency.
3. Develop a study guide that explains the principles and rewards of airmanship and stresses the importance of continuous learning and self-improvement.

OVERALL TEACHING STRATEGIES

1. Use a systematic approach to teach basic and higher-order skills. Early training should focus on psychomotor skills (manual flying), basic procedural skills (understanding of situations and procedures), and introduce the concepts needed later (e.g., systems knowledge). Intermediate training should refine psychomotor performance, complex procedural skills, and expose students to the range of less predictable situations that can arise and how they have been resolved in the past. Later training should focus on the development of higher-order skills such as problem solving.
2. Allow students simulator practise in between formal instruction to facilitate continuous learning, reinforce learning and aid retention.

3. Encourage students to keep a journal for recording minor errors during a flight and new techniques that were employed. Keeping a journal allows for a period of reflection, records tacit knowledge and provides direction for areas of improvement.

4. Provide examples of aviation problems that students may have to face or have already faced (problem-based learning).

5. Encourage *mental flexibility* by using multiple representations of training content using several kinds of media.

6. Avoid oversimplifying training content. Present a number of examples to make apparent, rather than hide, the variability and interconnections of concepts and themes. Knowledge should be highly interconnected rather than compartmentalised.

7. Use *cognitive apprenticeship* to teach cognitive skills. This approach borrows from the features of a traditional apprenticeship e.g. one-to-one teaching, but focuses on teaching cognitive skills. The role of the instructor is to make explicit the cognitive processes associated with problem solving and decision making, to teach tacit knowledge as well as textbook knowledge and to encourage students to try out different strategies and observe their effects.

8. Support students by performing parts of the task they cannot perform and gradually reduce the amount of "scaffolding", shifting more of the control to the learner.

9. Encourage students to give reasons for their actions making their tacit knowledge more explicit.

10. Encourage students to try out different strategies and observe their effects. If students develop misconceptions, confront them with anomalies and counter-examples.

11. Provide opportunities for collaborative learning to create a culture of teamwork and increase motivation.

12. Use CBT to teach problem-solving skills in order to make the thinking processes of the learner explicit. Tailor instruction to suit the learner via a pre-test.

TEACHING THE FOUNDATIONS OF AIRMANSHIP

1. Teach discipline and attitudes from the outset.

2. Teach students to identify the hazardous attitudes associated with poor airmanship and the various antidotes (see Table 2). Egotistical attitudes are the antithesis of the values embodied in airmanship. Airmanship values include *maturity* — being able to admit a mistake and reverse a decision if necessary, and taking personal *responsibility* for ensuring safe flight and meeting mission objectives.

Hazardous Attitude		Antidote
Anti-authority	'The regulations are for someone else'	'Follow the rules. They are usually right.'
Impulsivity	'I must act now, there's no time'	'Not so fast. Think first'
Invulnerability	'It won't happen to me'	'It could happen to me'
Macho	'I'll show you. I can do it'	'Taking chances is foolish.'
Resignation	'What's the use?'	'Never give up. There must be something I can do.'

Table 2: The antidotes to hazardous attitudes [8]

3. Use case-studies to provide a context for technical knowledge.

4. Teach teamwork skills using videos and voice recordings of real-world examples, stopping the video at key points and discussing team interactions.

5. Teach strategies for managing workload (see Table 3). Modern aircrew have to cope with unprecedented levels of information that have the potential to overwhelm them. However, aircrew can be taught to manage their workload and maximise their spare capacity. Recognising when aircrew are overloaded (or underloaded) and taking steps to restore the balance is an important part of airmanship.

Workload Management Strategies
1. Be aware of own and team member spare capacities.
2. Stabilise the aircraft and manoeuvre to a safe position before attempting to balance workload.
3. Establish task priorities and filter irrelevant information.
4. Put low priority tasks on hold.
5. Delegate tasks to others.
6. Take on tasks when others are overloaded.
7. Manage distractions – recognise when aircrew are being distracted and re-establish priorities.
8. Establish roles and responsibilities prior to each flight.
9. Anticipate and rehearse plans during periods of low workload (also prevents underload).

Table 3: Workload management strategies [5]

6. Teach the fundamentals of situation awareness [10], how to recognise the signs of lost SA and how to recover SA (see Table 4). Put simply, situational awareness can be defined as "knowing what is going on, so you can figure out what to do"[11]. It enables aircrew to stay mentally ahead of an aircraft and develop the foresight needed to detect and resolve problems before they have chance to fully develop. Incident reports are littered with examples of aircrew who have made confident decisions based on inaccurate or lost SA; aircrew need to be able to assess the reliability of their own SA (metacognition) and take steps to regain it once it is lost.

Indicators of Lost SA
1. Ambiguity or confusion - feeling you are missing something, a sense of uncertainty.
2. Fixation – channelled attention or preoccupation on one activity or event.
3. Reduced frequency or poor communications - when we start to lose SA we stop talking.
4. Failure to stay ahead of the aircraft – reacting exclusively to immediate concerns rather than preparing and anticipating future events.
5. Failure to meet targets during a mission e.g. ETAs.
6. Use of undocumented procedures or violation of a minimum.
7. Attempting to operate aircraft systems outside of known limitations.

Steps to Regain SA
1. Buy some time – manoeuvre the aircraft away from potential hazards.
2. Stabilise the aircraft.
3. Seek information – visual, aural, seat of the pants. Resolve any discrepancies and restore confidence levels.
4. Learn from the experience - consider what cues were available that could have prevented a loss of SA.

Table 4: Strategies for Managing SA [3]

ASSESSING AIRMANSHIP

1. Assess airmanship qualities at an appropriate stage of learning. Student confidence can be fragile and will be dented if evaluated on skills that are too advanced.

2. Encourage students to perform self-assessment — a technique that has been shown to significantly improve retention of information.

3. Use objective data to assess physical skills (e.g. bombing accuracy, flight profile accuracy), thus providing concrete feedback for the student.

4. Define and use performance standards for both physical and cognitive skills — this ensures consistency in student evaluation and informs students of their position on the performance ladder.

5. Use "think-aloud" protocols as a source of data for assessing cognitive skills.

6. Use "concept maps" to test students' understanding of the interconnections between knowledge elements.

7. Use the "secondary task" method for assessing spare capacity. This involves systematically loading students with increasing demands until the point is reached at which primary task performance breaks down. This provides an indication of the students' spare capacity. Systematically increasing levels of workload can also be used to practice workload management strategies. The technique should not be used for *ab initio* students.

8. Introduce events during simulator training as an objective means of assessing airmanship performance. Students score a "hit" if they demonstrate the required behaviour and a "miss" if they

fail to demonstrate the required behaviour. Events could include emergencies to monitor students' skill in priority allocation, unscheduled aircraft to test student lookout discipline and erroneous aircraft communications to test communication skills.

9. Evaluate airmanship by assessing the processes employed in a mission and not just the outcomes. This provides a more diagnostic view of student performance than simply outcome measures alone.

An integrated approach is needed for the training of airmanship — an approach which combines traditional "stick and rudder skills" with higher-order cognitive skills and attitudinal skills, and an approach which employs modern teaching strategies inspiring students to seek performance excellence and encouraging greater accountability for learning development.

CONCLUSION

Within the aviation community, there are many definitions and concepts of airmanship. It is our contention that airmanship is logically defined as a personal state that enables aircrew to exercise sound judgement, uncompromising flight discipline and skilful control of an aircraft within a situation.

There are several indicators of effective airmanship. Primarily a good airman is someone who prevents things from going wrong by maintaining situational awareness, using foresight to anticipate potential hazards and making sound plans that take them into account. However, it is perhaps easiest to see effective airmanship in action when problems arise. Then, effective aircrew correctly assess the dynamic situation, apply sound judgement and take decisive and appropriate action. In doing so, aircrew manage the available resources and prioritise their goals; demonstrate good discipline and strong teamwork, communicate clearly to all appropriate agencies, and employ comprehensive knowledge and considerable expertise to the situation.

Whilst it is true that airmanship is a personal quality, and that individuals do vary in their natural ability for it, aircrew can be trained and motivated to achieve airmanship excellence through application of the Foundations of Airmanship model. Airmanship training centres on the building of three key skills of judgement, control and discipline using a defined series of knowledge and skills. However, airmanship is more than having the requisite skills and knowledge; it is about having an appropriate attitude and a desire to perform optimally at all times— this personal conviction will enable aircrew to attain the highest levels of airmanship performance.

Outstanding airmanship can only be achieved by placing airmanship training at the very heart of a training system rather than as an adjunct. An holistic approach to airmanship training is essential to enable aircrew to meet the new and significant challenges of modern aviation.

REFERENCES

[1] Training Development and Support Unit Flying Training Development Wing, Airmanship TDSYU/779/1/5/TRG 27 June 2000

[2] Hayes, T. (2002)Airmanship & flight discipline. http://www.auf.asn.au

[3] Kern, T. (1996). Redefining airmanship. McGraw-Hill.

[4] McGuinness, B. (2002) Technical note on airmanship.

[5] Ebbage, J. L. (2002) Understanding Airmanship. BAE SYSTEMS Advanced Technology Centre Report JS 14862

[6] Air Line Pilots Association http://www.alpa.org

[7] Federal Aviation Administration, Advisory Circular Aeronautical Decision Making (AC no: 60-22). Dec 13 1991.

[8] Jensen, R. S. (1995) Pilot Judgement and Crew Resource Management. Avebury Aviation.

[9] Fowlkes, J.E, Lane., N.E., Salas, E., Franz, T. and Oser, R.L. (1994) Improving the measurement of team performance: The TARGETs methodology. *Military Psychology*, 6, 47-61.

[10] McGuinness, B. and Foy, L. 'A subjective measure of SA: the Crew Awareness Rating Scale (CARS).' Paper presented at the *Human Performance, Situational Awareness and Automation Conference*, Savannah, Georgia, 16-19 Oct 2000.

[11] Adams, E.C. (1993) Fighter cockpits of the future. Proceedings of 12th DASC, the 1993 IEEE/AIAA Digital Avionics Systems Conference, 318-323.

Correlation Between The System Of Values And The Stressful Factors (Anxiety Level) Among The Group Of Georgian Soldiers

Maj. Z. Avazashvili, M.D., Ph.D.
Central Military Hospital
6 Navtlugi St, 380079 Tbilisi
Georgia
E-mail: avaza@gol.ge

T. Chanturishvili, M.D., Ph.D.
2/2, Chavchavadze Ave
380079 Tbilisi,
Georgia
E-mail: chanturishvili@land.ru

T. Bjalava, Ph.D. in Psychology
Tbilisi State University
1, Chavchavadze Ave
380079 Tbilisi,
Georgia

S. Chanturishvili
Tbilisi State University
1, Chavchavadze Ave
380028 Tbilisi,
Georgia

T.Morbedadze
Tbilisi State University
1, Chavchavadze Ave
380028 Tbilisi
Georgia

Summary

The anxiety level and the system of values have been studied with help of the situational anxiety test (V.Norakidze, in I) and the scale of values (Rokich, in 2) among 180 Georgian soldiers. As a basis to this investigation there served a preliminary supposition that social problems, existing in the country, experiencing hardships at the early stages of its independence, would reveal very high anxiety level and a motley spectrum of the scale of values among the soldiers to be investigated. The results of the study showed an absence of the extreme high anxiety level. They also showed a system of values, not very relevant to military persons – preferences to those ones, which (as the examined persons thought) would guarantee for them a comfortable life: love, good friends, happy family life. On the contrary, such values, as discipline, active and productive life, serving and the like were disregarded.

Introduction

While planning this investigation we intended to find out what are the psychosocial and psycho-individual peculiarities of soldiers, how do they adjust to contradictions and difficulties of today's Georgian life. Obviously, there are too much of them: total corruption - the legacy of completely viciously organized social relations (of which the country is not yet liberated), independence of the native land – wretched mockery of such a high notion, fundamental concepts experiencing revision, principles, settled long ago going to ruin, social psychics under hard blows etc. Studying the psychological state of soldiers against such a background is not only of sheer interest, but of great actuality too (3); in order to decrease the feeling of discomfort, we ourselves find comfort in our aspirations and actions (4). It can be sad, that for lessening unpleasant senses we, as a rule, adjust our thinking to being and try to regulate our thoughts as much as possible.

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

Methods

The study has been carried out on 180 soldiers of 17 – 25 years age from 3 different military units, or 3 groups: I – soldiers, undergoing medical treatment in a military hospital, II – students of the Military Academy, III – soldiers of an ordinary military unit.

Because of identical results among the investigated persons it was possible to unite all mentioned groups in one. Results were counted with help of SPSS method and happened to be statistically reliable ($P<0,05$).

Anxiety levels have been studied with help of 2 tests in form of closed questionnaire (1). One of them consists of 60 objectives to be answered with “yes” or “no”. Relatively, different questions summarize in scores differently. According to this questionnaire, 5 levels of anxiety can be distinguished:

Number of scores	**Levels of anxiety**
40 – 50	**Very high**
25 – 40	**High**
15 – 25	**Middle, tending to high**
5 - 15	**Middle, tending to low**
0 – 5	**Low**

Table 1: Number of scores and levels of anxiety

The second test consists of 30 objectives, which need answers: “never”, “rarely”, “sometimes”, “mostly”, and “always”. These answers have been numbered from 1 to 5 and the final amounts of scores have been related to one of three types of individuals:

Number of scores	**Individuals**
100 – 150	**A, stressful persons**
76 – 99	**AB, intermediate persons**
30 – 75	**B, quiet persons**

Table 2: Number of scores and characters of individuals

A system of values has been investigated with help of the scale of Rokich (2). According to it, the values of two types were picked out – the terminal ones, i.e., the values-goals (18 ones) and the instrumental ones, i.e., the values-means (17 ones). It is clear, that the terminal values are such notions, which are the final goals of individual’s aspirations. As to instrumental values, they are the means, by which the terminal values can be achieved.

Each of these values has been written on a separate card and soldiers have been asked to range them – they had to lay out separately the cards with the values that they preferred. The ignored values created the group of the neglected ones.

Results

The results for the anxiety level, tested with the questionnaire, consisting of 60 questions occurred to be situated in limits of 5 – 40 score (middle – high levels of anxiety, the most of them belonging to the group "Middle, tending to high"). According to second anxiety questionnaire, the results show the intermediate sum of scores, 76 – 99.

	Group 1	Group 2	Group 3	All
Test 1	**22,47**	**20,58**	**21,47**	**21,16**
Test 2	**86,15**	**84,12**	**82,11**	**84,01**

Table 3: Average results for all groups

No reliable difference exists between the data of 3 groups ($p>0,05$)

As to the scale of values, the soldiers preferred such instrumental values, as good breeding, honesty, strong will and such terminal values, as health, happy family life, love, good friends.

The whole spectrum of preferred and neglected values is shown in the tables 4 and 5.

Preferred		Neglected
2,1%	Active and effective life	**7,1%**
2,4%	Vital wisdom	**3,3%**
22,8%	Health	**2,1%**
5,7%	Interesting job	**1,9%**
1,6%	Beauty of nature and of art	**6,7%**
11,5%	Love	**1,6%**
2,2%	Material well-being	**6,6%**
11,3%	Kind and devoted friends	**0,3%**
2,0%	Social acknowledgment	**5,7%**
1,8%	Cognition	**8,9%**
1,4%	Productive life	**9,2%**
2,7%	Development (physical and spiritual)	**4,1%**
1,7%	Entertainment	**5,4%**
5,7%	Freedom (of judgment and action)	**5,6%**
19,6%	Happy family life	**2,0%**
2,9%	Happiness of others	**7,0%**
1,0%	Creative activity	**13,3%**
1,7%	Self-confidence	**6,9%**

Table 4: Preferred and neglected terminal values.
[Because of identity of results, all three groups are united].

Preferred		Neglected
4,5%	Accuracy	**3,4%**
22,0%	Good breeding	**2,0%**
11,1%	Cheerfulness	**2,3%**
3,8%	Discipline	**3,6%**
3,9%	Independent behavior	**4,6%**
2,1%	Irreconcilability to own and others shortcomings	**15,6%**
8,8%	Education	**1,7%**
4,0%	Responsibility	**3,3%**
2,5%	Rationalism	**13,4%**
3,5%	Self-control	**2,6%**
3,5%	Defense of own views	**2,4%**
5,7%	Strong will	**1,8%**
1,9%	Conciliation	**12,7%**
1,9%	Width of views	**16,1%**
18,5%	Honesty	**6,6%**
1,0%	Productive work	**8,9%**
2,8%	Thoughtfulness	**2,1%**

Table 5: Preferred and neglected instrumental values.
[Groups are united as in table 4].

Discussion

What are the data of our investigations and how can they be interpreted? Is there anything remarkable in them? We consider, it is.

The data of the investigated anxiety level show mainly middle, tending to high level of it. Perhaps, the absence of deviation toward the extreme high degree is a sign of adaptation to existing circumstances. It is impossible to live permanently in conditions of high anxiety. This changes profoundly our physiology and psychology and can not go on for a long time. Therefore, the limits of the anxiety levels in our study seem to us logical.

Very interesting, by our opinion, are the results of study of social values. Both terminal and instrumental values are the figures of person's psychological adaptation to the environment. The kind of adaptation is a result of the structural characteristic changes of a person. Each individual, with his or her own particular requirements, psychic experiences and functions, has personal attitude to the environment. This environment, influencing on the person and provoking his or her modification, creates individual's actual disposition. A collection of preferred and neglected or disregarded values in our study do reveal such a disposition, that seems not fully relevant to the military persons, especially to those, studying at a Military Academy. It is difficult to consider and estimate each value, all the nuances of it, but neglectfulness of active and effective job, productive life, self-confidence, accuracy and discipline seems rather strange. It is not clear, why such value, as "material well – being" was neglected.

A portrait of the soldier, revealed by our study, as we think, is a result of the reality, in which our country has been existing for last 200 years; it has gradually lost the experience of being a state. The homeland was the compounding part of the immense empire, which could not be identified with a small country, needing to be served.

False, utopian values, imposed on people and not stimulating patriotic feelings, have also negatively influenced on their psyche.

Our investigation demonstrated the promotion of values, which – as young people think – defend them morally, create a comfortable surrounding for them (3). On the contrary, some fundamental principles, based on serving, discipline and responsibility are neglected.

Disclaimer

The findings, views and opinions in this report are those of the authors and should not be construed as the official Army position, policy, or decision unless so designated by other official designation. Soldiers voluntarily participated in those studies being informed on the general idea of the study. Approved for public release; distribution unlimited.

References

1. R.Rimskaja, S.Rimsky – Psychological tests, or how to know oneself. Moscow, Ast-press book, 2003 (in Russian).
2. Psychological tests – vol.I. Edit.: A.A,Karenina, publish. center "Vlada", Moscow, 2002 (in Russian).
3. Sh.Nadirashvili – Social psychology of a person. Tbilisi, University edition, 1975 (in Georgian).
4. D.G. Myers – Social Psychology, 1996.

Training Critical Thinking for Tactical Command

Dr. Karel van den Bosch
Drs. Anne S. Helsdingen
Drs. Marlous M. de Beer

TNO Human Factors
PO Box 23
3769 ZG Soesterberg
the Netherlands

tel: +31 346 356211
fax: +31 346 353977
email: vandenbosch@tm.tno.nl

Military missions are more and more focused upon peace enforcing operations in regional conflicts. There is often uncertainty about the intentions, capabilities and strategies of the parties involved. Successful operations under such unstable and complex conditions require competent commanders and staff personnel.

Recent studies have shown that experts in military tactical command treat decision making as a problem-solving process. Experts have large collections of schemas, enabling them to recognise a large number of situations as familiar. Another capacity of experts are their problem solving skills if an immediate match between the actual problem situation and available schemas in memory cannot be established. When faced with an unfamiliar tactical problem, experts collect and critically evaluate the available evidence, seek for consistency, and test assumptions underlying an assessment. They then integrate results in a comprehensive story. This expert's approach has been used to develop critical thinking training.

This paper presents empirical studies into the effects of critical thinking training. Individual commanders and commanding teams played scenario-based exercises in both simplified and high-fidelity task environments. Half of the participants received instruction, guidance, and feedback in critical thinking. The other half received the same scenarios, but without specific support. After training, test scenarios were administered to all groups. Results showed positive effects on the *process* of tactical command (i.e. better argumentation for situation assessment) as well as on the *outcomes* (i.e. more and better contingency plans). In addition, members of critical thinking training teams were more inclined to clarify their perspective of the situation to each other.

Critical thinking training supports commanders in situation assessment and decision making, and stimulates team members to engage in activities required to develop a shared mental model and to co-ordinate actions. For the Netherlands Navy and Army, the results of the studies are now being used to implement critical thinking into tactical decision games in order to develop new military training programs that will enhance the acquisition of sophisticated domain knowledge and decision making skills.

1. Introduction

In recent decades changes in the (inter-) national political situation has demanded a reflection on the way military personnel have to prepare for military missions. Peace keeping and peace enforcing missions have become more varied, complex and unpredictable. These missions have to be fulfilled with fewer personnel, in many different circumstances. At the same time, the professional development of our military officers has shifted towards more civil areas such as business administration, public administration, socio-economic studies, logistics and humanities. But with the number of deployments increasing, combat capabilities and skill in military tactics regain popularity. Training and education programs have to be adapted to provide military personnel and especially military commanders with training aimed at the successful preparation, execution and management of military operations in unstable and complex conditions (Lussier, 2003).

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held in Genoa, Italy, 13 – 15 October 2003, and published in RTO-MP-HFM-101.

Studies into expert strategies in tactical decision making have shown that experts have large collections of schemas, enabling them to recognise a large number of situations as familiar. Furthermore, when faced with novel situations, experts apply deliberate problem solving strategies that differ significantly from those of novices. When faced with a complex and unfamiliar tactical problem, experts collect and critically evaluate the available evidence, seek for consistency, and test assumptions underlying an assessment. They thus try to integrate the results in a comprehensive, plausible, and consistent story that can explain the actual problem situation, whereas novices very often consider aspects of the situation separately and independently. The experts' approach is used to develop a new training concept: *critical thinking* (CT) (Cohen, & Freeman, 1997; Cohen, Freeman, & Thompson, 1997; Cohen, Freeman, & Thompson, 1998).

The critical thinking strategy involves a problem solving approach to new and unfamiliar situations. It is a highly dynamic and iterative strategy, consisting of a moderately sized set of methods to build, test and critique situation assessments. These methods are to some extent generalizable but they can only be taught if grounded in a specific domain and trainees have already a certain level of knowledge of that domain. Effective critical thinking training combines instruction with realistic practice (Cohen, Freeman, & Thompson, 1998). Practice in the form of scenario-based training is considered an appropriate approach to training competencies required in complex task environments (e.g. Fowlkes, Dwyer, Oser & Salas, 1998; Oser, 1999). The design of exercise scenarios is very important since these have to provide opportunities to practise critical thinking strategies. Some guidelines for the design and execution of scenarios for critical thinking training are described in the next paragraph.

The effects of the critical thinking approach to tactical command and control training have been studied in several explorative studies (e.g. Cohen & Freeman, 1997; Klein, McCloskey, Pliske, & Schmitt, 1997). These studies showed promising results but they were conducted with individual decision makers, in simplified training environments and performance of critical thinking trainees were compared to trainees that did not receive any training. We conducted two training studies in which we compared performance of our critical thinking group with the performance of a control group that conducted the exercise scenarios with standard instruction and feedback. In the first study we focussed on the individual military decision maker and we conducted the training in a very simplified training environment. In the second training experiment we studied teams in more complex and dynamic training environments. These studies are described in paragraph 3.
In paragraph 4 we will briefly outline the approach to our implementation study of critical thinking training. This study is still being undertaken, so only some preliminary findings can be reported.

2. Scenario generation and instruction methodology for critical thinking training

In scenario based training, trainees prepare, execute and evaluate exercises that are simplified situations of the real world. A scenario has a starting point and depending on the type of scenario, events are specified in time and space. Scenarios can be very structured in the sense that all events are scripted, or have a free play character. Also, scenarios can differ in complexity: they may be simplified, leaving out many aspects of the real world, or they are complex, incorporating many aspects of the operational task environment.

Scenario based training provides trainees with the opportunity to build domain specific experience under controlled and safe conditions (Farmer, van Rooij, Riemersma, Jorna, & Moraal, 1999). By executing training scenarios, trainees may gain knowledge about typical problems and their solutions, thereby increasing their experience database of situation-response relationships. For critical thinking training, it is important that the training scenarios provide trainees with the opportunity to practise critical thinking skills. These involve (Helsdingen & van den Bosch, 1999):

<u>Creating a story</u>: A story is a comprehensive assessment of the situation, in which all the existing evidence is incorporated and explained and assumptions are made about uncertain aspects of the situation. Past, present and future are addressed in the story. The purpose of story building is to keep trainees from assessing

situations solely on isolated events. Instead, trainees are taught how they can integrate the available information into its context, which may include elements as the history of events leading to the current situation, the presumed goals and capacities of the enemy, the opportunities of the enemy, etc.

Testing a story: Testing a story is aimed at identifying incomplete and conflicting information. They have to correct these problems by collecting more data, retrieve knowledge from memory, or make assumptions about the missing piece of the story or to resolve conflicts in the argumentation.

Evaluating a story: After a story is constructed, it should be evaluated for its plausibility. The decision maker has to take a step back, identify critical or hidden assumptions and play the devils' advocate by falsifying these assumptions, i.e. explaining how this assumption can be false and building an alternative story.

Time management or the Quick test: Critical thinking is not always appropriate. Decision makers have to evaluate the time available and the consequences of their actions. In stressful situations such as those often encountered by military commanders, usually there is little time to spare. The decision maker should act immediately unless the risk of a delay is acceptable, the cost of an error is high, and the situation is non-routine or problematic (Cohen, Freeman, & Thompson, 1998). Critical thinking training focuses on the way trainees apply these criteria.

Our guidelines for scenario development follow directly from these critical thinking skills. An exercise scenario can be based on a real world course of events; however, some adjustments have to be made.

1. The real world course of events is probably too complex, with too many factors playing a role. It is important to simplify the scenario, especially for initial training exercises. Domain experts can identify critical factors and some distracting evidence.

2. The scenario should be built in the same way that trainees have to build their story. The developer and domain expert have to develop a coherent story, identify critical assumptions and come up with alternative explanations. This will help to decide what information has to be presented to the trainee and what uncertainties or conflicting evidence have to be introduced into the scenario.

3. The scenario developer or instructor has to consider the strategies by which the trainees can correct their stories. Do they have the knowledge to retrieve extra information from their memories; can they collect additional information, what kind of assumption could they come up with to resolve the gaps or conflicts in the information? This consideration will help the instructor or scenario players to anticipate on trainees' questions or behaviour during the actual exercise.

4. Exercise scenarios have to be varied and challenging. This means that the domain specific problems have to be challenging and varied, in order to prevent that critical thinking strategies become a trick.

5. What and when trainee performance has to be measured is something that has to be identified during the scenario development. Performance measures should be aimed at the outcome and processes of task performance. The outcome measures such as situation reports, orders, plans, and contingency plans are domain specific and linked to particular scenario events. They have to be designed and evaluated by domain experts. The process measures have a more generic character and they refer directly to the critical thinking skills. They can only be evaluated by domain experts, within the context of a specific scenario. Process measures include: information processing (selecting relevant information, story building, identification of incomplete or conflicting information), argumentation (the explanations for missing or conflicting evidence, criticising assumptions, coming up with alternative explanations), time management skills (make efficient use of the available time), and team skills (communication, supportive behaviour, co-ordination, leadership).

6. Test scenarios have to be developed for the measurement of transfer of training. Performance on the test scenarios has to be evaluated by independent experts, to prevent biased judgements by the familiar instructor. For the interest of evaluation of training effectiveness the independent expert should also be blind to the experimental manipulations.

3. Training studies

STUDY 1

The first study is conducted in the domain of "air defence" of the Royal Netherlands Air Force, in particular the Tactical Command Station (TCS) of a ground-to-air defence battalion. In an office room, trainee-officers played air-defence scenarios under supervision of a scenario leader. The trainee played the role of battle captain, the scenario leader played all other functions (lower and higher control), and introduced the scripted events in the scenario (e.g. battle damage reports, information about enemy movements, identified radar tracks). Prior to each training scenario, the trainee was provided with a short description of the political, military, and civil background situation.

Ambiguous, incomplete and inconsistent information was intentionally introduced into the scenarios to allow for alternative interpretations of events.

METHOD

Design: A training-posttest design was used (see Table 1). The supervising project officers arranged participants according to their tactical education and experience, and assigned matched pairs of trainees randomly to conditions.
Briefing and instruction: The critical thinking group received a critical thinking tutorial, followed by a demonstration in which two scenario leaders (one of them played the role of trainee) showed how critical thinking should be used in the scenarios. Trainees of the control group were instructed to run the scenarios as a normal command post exercise.
Training general: Two sets of three scenarios were used. Two scenario leaders were available. Order of scenario sets, and assignment of sets to scenario leaders was balanced. While performing the scenarios, trainees were asked to think aloud to give the scenario leader access to the assumptions and reasoning underlying the assessments and decisions. At pre-specified moments, the scenario leader "froze" the scenario for interventions (see below). After each scenario, the scenario leader filled in an evaluation form.

Critical Thinking group: critical-thinking supporting schemes were available during training. At scenario freezes and after completing the scenario, the scenario leader provided support and feedback on the critical thinking process (e.g. by asking "which alternative explanations are possible?" or "how can you verify that assumption?").

Control Group: trainees received outcome feedback only (e.g. "that was a good decision", or "you should have issued that request earlier").
Test: Two test scenarios and two scenario leaders were available. Order of scenario and assignment of scenario to scenario leader was balanced. All trainees were asked to think aloud. No support or feedback was given.

Condition	Instruction	Training	Test
Critical Thinking-group (N=8)	Instruction and demos in critical thinking	Scenarios 1-6, with support in critical thinking process- and outcome feedback	Scenarios 7-8; without support; no feedback
Control-group (N=8)	No specific instruction	Scenarios 1-6, no support outcome feedback only	Scenarios 7-8; without support; no feedback

Table 1: Research design

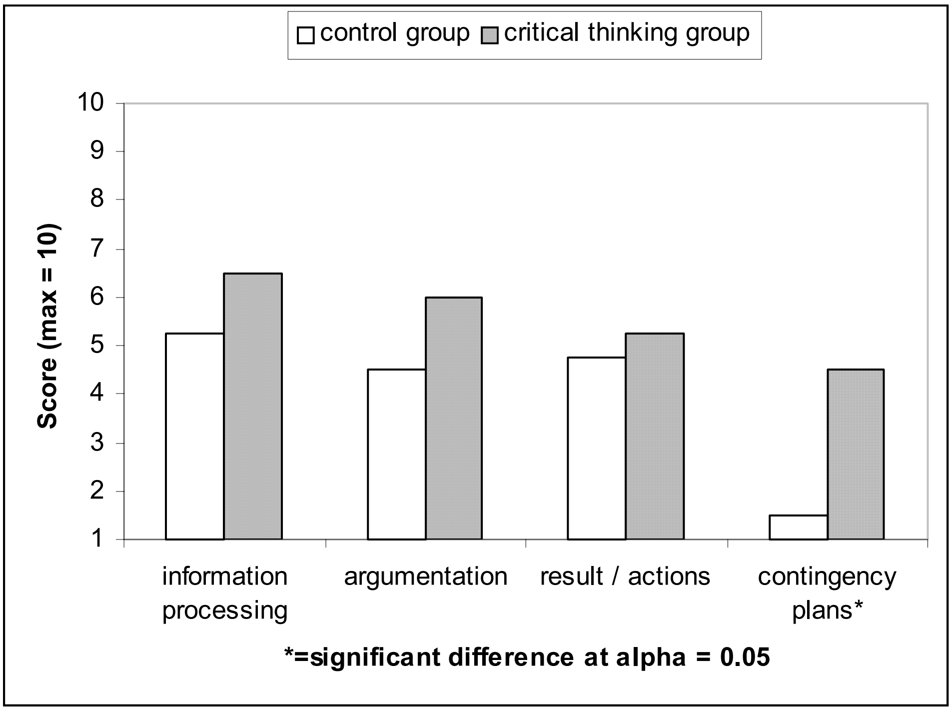
Performance measures: outcome measures were used to assess the quantity and quality of the end result (what is actually achieved?); process measures to describe the strategies, steps or procedures used to accomplish the task. Result and contingency plans were used as outcome measures; information processing and argumentation as process measures. Scenario leaders evaluated trainee performance on these variables on a 10-point scale. A verbal description was used for each scale point, ranging from "very poor" for score 1, to "excellent" for score 10. Prior to the experiment proper, scenario leaders had used the results of a pilot-study (using the same scenarios but with different trainees) to come to a common understanding of assigning scores.

Statistical analysis
Since the performance measures are ordinal data, we had to perform non-parametric statistical analysis. Data were analysed by means of the Kruskal-Wallis test.

RESULTS

Figure 1 shows the median scores on the test scenarios. The critical thinking group seemed to perform slightly better than the control group on all variables.

Significant differences between groups were only found for contingency plans ($H(1)=3.91$, $p<0.05$). The variables information processing, argumentation and result showed a similar pattern, but the differences between groups were not significant ($H(1)=1.62$, $p=0.21$; $H(1)=2.08$, $p=0.15$; and $H(1)=1.23$, $p=0.27$, respectively).



/

Figure 1: median scores on the test scenarios

DISCUSSION

Trainees of the critical thinking group performed a little better than the control group. This indicates that the critical thinking approach might be a sufficient tool for improving the quality of tactical command. The results of the study had practical implications for the military training organisation as well. The scenario leaders discovered that this form of training disclosed gaps in tactical education of the participating officers that had remained concealed in the large-scale exercises constituting normal training.
The outcomes corroborate the positive outcomes of earlier explorative studies (e.g. Cohen & Freeman, 1997; Klein et al., 1997). However, further research is needed to investigate a number of questions. First, scenario

leaders' scoring may have been biased by their knowledge of the training intervention and the design of the study. More independent assessments are needed. Second, the effect of training has been studied in a simplified task environment. Eventually, critical thinking skills need to be applied in the real world. For reasons of transfer it is necessary to investigate whether critical thinking skills can be successfully trained in high-fidelity task environments. Finally, the training of the first study focused upon the individual commander, whereas tactical command is typically performed in a team. The effects of critical thinking training for teams need to be determined. These questions are addressed in study 2.

STUDY 2

This study investigates the effects of critical thinking applied to the training of command teams operating in their natural task environments. It is conducted in the domains of "anti air warfare" (AAW) and "anti surface warfare" (ASuW) at the Operational School of the Royal Netherlands Navy. Teams of trainees played single ship / single threat scenarios in a high-fidelity tactical simulator. The ASuW and AAW teams consisted of an officer and petty officer.

METHOD

Design: A training-posttest design was used. The supervising project officer arranged the eight participating teams according to their tactical education and operational experience, and assigned matched pairs of teams randomly to either the "critical thinking training" group or the "control" group. The supervising project officer selected two instructors for the study. They were randomly assigned to conditions.

Briefing and instruction: Prior to the experiment proper, instructors assigned to train "critical thinking training" teams were extensively briefed on the critical thinking training method, as well as how to support trainees in the application of critical thinking processes. Instructors assigned to the control team were not informed about the concept of critical thinking. They were told to support the teams as they would normally do in training. Instructors trained one team at a time. The briefing, training and testing required four days per team.
The first day of the actual study was used for briefing and instruction of the teams. The experimenter and the assigned instructor briefed the "critical thinking" team on the principles of critical thinking and showed them how to apply this principle in paper-based demonstration scenarios. The control group instructor briefed his team on the itinerary of the coming days, and discussed a paper-based demonstration scenario with them.

<u>Training general</u>: On the second day, teams received two interactive role-playing scenarios in a staff room under supervision of their instructor. On the third day, teams received two scenarios in the tactical simulator. A scenario run took approximately two hours. See study 1 for details on how the instructor made interventions to support learning.
Critical Thinking group: The instructor encouraged his team to explicitly execute all critical thinking components and he provided extensive guidance and feedback during and after the scenarios.
Control Group: The instructor supported the control group teams as in normal training.
<u>Test</u>: On the fourth and final day, teams were tested on two test scenarios in the simulator. Instructors were not present. Two independent subject matter experts evaluated the performance of trainees individually, as well as that of the team. They received the scenarios on paper. Markers in the scenario description prompted the evaluators to score trainee and team performance at that particular moment, on specified performance criteria.
Evaluators were not informed about the concept of training nor of the purpose and the design of the study. Assignment of evaluators to teams and to scenarios was balanced.

Performance measures: the same outcome and process measures as in study 1 were used. In addition, performance with respect to time management and team skills was also scored. Because the evaluators were used to using the official NATO 4-point scale, it was decided to use this 4-point scale in this study as well.

The verbal descriptions for the four scale points are, respectively: 1= UNsatisfactory, 2= MArginal, 3= SAtisfactory, and 4= EXcellent.
Prior to the experiment proper, the experimenter briefed the evaluators about the scoring procedure and how to use the scale. The results of a pilot-study were used to arrive at a common interpretation of performance measurement.

Statistical analysis
Since the performance measures are ordinal data, we had to perform non-parametric statistical analysis. Data were analysed by means of the Kruskal-Wallis test.

RESULTS

Data on individual as well as on team performance were collected during training and test. For reasons of brevity, performance data of the two test scenarios will be reported only. Figure 2 shows the results on the test scenarios.

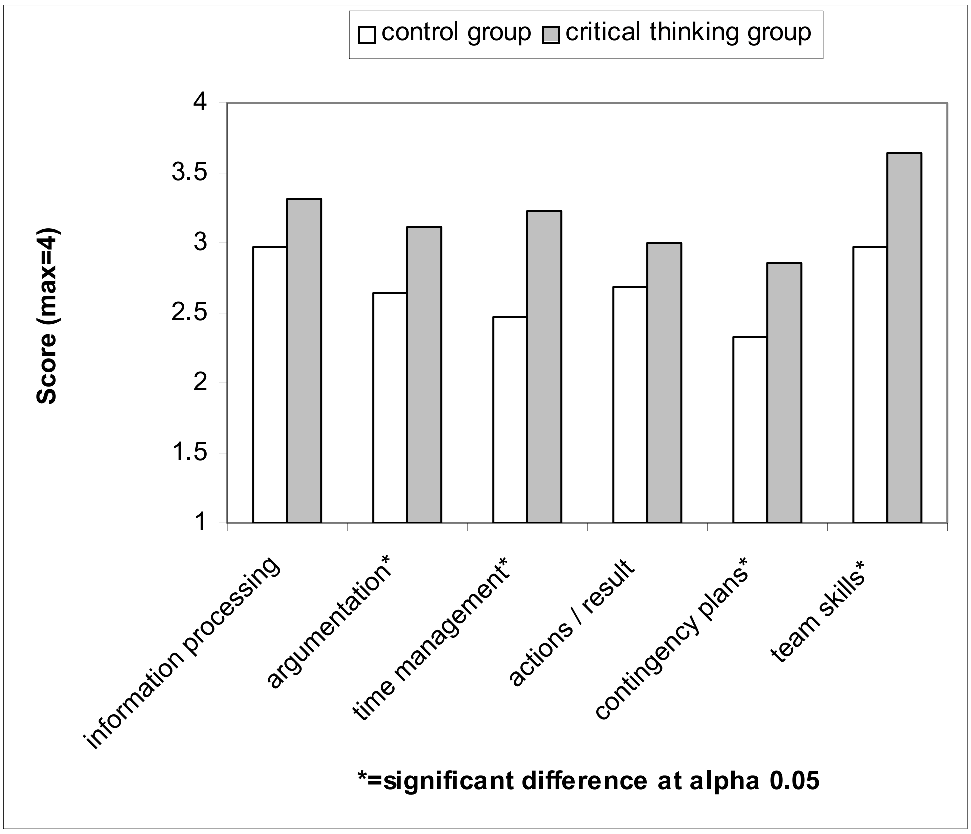


Figure 2: median results on the test scenarios

Univariate tests showed significant differences for argumentation ($H(1)=7.5$, $p<0.05$), time management ($H(1)=11.4$, $p<0.05$), contingency plans ($H1)=5.6$, $p<0.05$), and team work ($H(1)=8.7$, $p<0.05$). Performances on information processing and actions were not significantly different between groups.

GENERAL DISCUSSION

Critical thinking training produced positive effects on the process of tactical command (i.e. better argumentation for situation assessment) as well as on the outcomes (i.e. more and better contingency plans). The method supports not only individual commanders in situation assessment and decision making, it is also

particularly suitable for team members to clarify their assumptions and perspectives on the situation to the other team member(s). This is especially important for developing shared mental models and to co-ordinate team actions (Stout, Cannon-Bowers, & Salas, 1996). The dynamic and interactive nature of high-fidelity simulator training sometimes provides too little opportunity for the object of this type of training: reflective and critical task performance. This can be overcome by preparatory paper-and-pencil- and role-playing scenarios, and by introducing pauses in the simulator-scenarios.
The present and earlier studies (Cohen, & Freeman, 1997; Cohen et al., 1997; Cohen et al., 1998; Freeman & Cohen, 1996) now provide sufficient evidence warranting the implementation of this type of training in practical (military) training programs. Such implementation studies are needed to provide answers to important questions, like: "how can we integrate critical thinking training into an existing curriculum?", "what instruction and training do observer/trainers need for successful application (train the trainers)?", "what is the transfer of training?", and "what are the long-term effects?"

4. Implementation of critical thinking training

Our training studies have shown positive results for critical thinking training. However, these training studies were specifically designed to study the effects of our training manipulation. They were mini-training courses, conducted under controlled experimental conditions. This is very different from any normal training program. In a regular training program, instructors usually have the freedom to treat every student differently, according to the individual students' needs. In our training experiment, instructors had to treat all students similar and according to specific rules for reasons of standardisation. Another difference is the short duration of a training study as compared to a regular training program. A short training intervention may not provide the opportunity for trainees to have mastered critical thinking skills.

The Operational school of the Royal Netherlands Navy has recently revised the training program for CIC commanders. Two separate curricula, one for the air warfare commander and one for the surface and subsurface commanders, are merged into one. This demanded a review of the classes and teaching materials. One of the students' comments on the training program concerned the (perceived) lack of correspondence between classroom instruction on the theoretical basics of naval warfare and the practical exercises in the tactical trainers. As part of an effort to introduce an integrated teaching method for theoretical education and practice, we have started an implementation study for scenario-based critical thinking training.

APPROACH

The implementation study involved many work packages, such as:

- the investigation and evaluation of the current training program,
- selecting a suitable course,
- identification of the appropriate moments for a training intervention,
- design of instruction materials for instructors and trainees,
- conducting try-outs with domain experts,
- the actual implementation of the critical thinking module.
- evaluation of the training: qualitative (interviews) and quantitative (performance measurement on standard test scenarios)

The CIC commander training program consists of several separate courses. Our implementation study focused on one of these courses, i.e. Anti Surface Warfare (ASuW). Within this course, trainees have to attend classroom sessions on theoretical background and principles of ASuW and apply this knowledge in practical exercises in the tactical trainer. The scenario-based critical thinking training aimed to facilitate this step by introducing tactical decision games (i.e. paper-and-pencil scenarios). The scenarios for these exercises were developed by experts of the operational school, according to the guidelines set out in this paper.

Prior to the critical thinking training we have instructed the scenario leaders and instructors extensively on the background and principles of critical thinking. We developed readers for the trainees on critical thinking training within the context of surface warfare, and organised a classroom instruction on critical thinking for trainees. Observation protocols and performance measures were designed to support instructors in their tasks.

Critical thinking training consisted of 4 sessions of 4 hours. The first session was the classroom instruction, the other three sessions were scenario based exercises. In each session one scenario was discussed. Trainees were A scenario consisted of a description of a situation. Trainees were assigned a specific role in this scenario (e.g. ASuW commander or SAG commander) and had to develop one or more plans. Trainees were encouraged explicitly execute all critical thinking components. In the first two scenarios a trainer/scenario leader provided extensive guidance and feedback during and after the scenarios, in the last scenario it was expected that teams guided themselves through the critical thinking strategy.

During the exercises, scenario leaders evaluated the teams' critical thinking processes and their resulting plans. For a description of the performance measures see the training studies. After the last training session, we asked trainees to fill out an evaluation form and provide us with their opinion on critical thinking training. Until now, the final evaluation has not been undertaken.

PRELIMINARY FINDINGS

The students were enthusiastic and motivated to co-operate. They found that critical thinking helped them reasonably well to systematically assess a situation, integrate different observations into a coherent story, identify uncertainties and justify assumptions, and come up with (contingency) plans. They appreciated the simple scenario based exercises as a suitable method for applying their tactical knowledge and practising their skills in tactical decision making. Although some trainees did express that they were not completely convinced about the surplus value of the critical thinking approach when taken into account that it takes valuable time. Our impression was that some students did not have enough domain knowledge to incorporate the strategy and apply it as a context-dependent problem solving strategy, instead they treated it as an independent and obligatory thinking procedure. Very often, these students did not even recognise the uncertainties, ambiguousness or conflicts in the scenario. Even during after action reviews they labelled many alternative situation assessments as improbable.

Critical thinking in teams is a fruitful strategy to develop a shared understanding of the situation, to avoid misunderstandings, and to develop contingency plans since criticising one another's assumptions, playing the devils advocate, and coming up with alternative explanations is easy.

The instruction and training for the instructors and scenario leaders is a critical factor in the implementation of a training module in a training program. We found that the instructors of the Operational School could not conduct the critical thinking training exercises without our constant support. This may be due to insufficient time to train and prepare the instructors.

We can conclude that the critical thinking approach is a suitable strategy for scenario based training in tactical command. The critical thinking approach encompasses guidelines for the design of effective training scenarios and instruction and support for reflective decision making. It is especially helpful when introduced in a simple training environment, with sufficient time and support to practise all components of the strategy. There are a few important considerations when implementing this training strategy. Instructors and scenario designers should be extensively prepared and instructed into the principles and methods of critical thinking. Furthermore, the approach only works within the framework of a specific domain and trainees should have a sufficient level of knowledge of that domain in order to reflect critically on their decision making process.

The long-term effects of critical thinking training are not yet determined. We hope to be able to conduct our final evaluations in the near future, and maybe conduct a follow up study to monitor task behaviour of our trainees in their future work environment.

5. References

Cohen, M. S., & Freeman, J. T. (1997). Improving critical thinking. In: R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: emerging themes and applications* (pp. 161-169). Brookfield, Vermont: Ashgate.

Cohen, M. S., Freeman, J. T., & Thompson, B. B. (1998). Critical thinking skills in tactical decision making: a model and a training strategy. In: J. A. Cannon-Bowers, & E. Salas. *Making decision under stress: implications for individual and team training* (pp. 155-190). Washington, DC: APA.

Cohen, M. S., Freeman, J. T., & Thompson, B. B. (1997). Training the naturalistic decision maker. In: C. Zsambok, & G. Klein. *Naturalistic decision making*. Mahwah, NJ: Erlbaum.

Farmer, E.W., Rooij, J.C.G.M. van, Riemersma, J.B.J., Jorna, P.G.A.M., & Moraal, J. (1999). Handbook of simulator –based training. Aldershot, UK: Ashgate Publishing Ltd.

Freeman, J. T., & Cohen, M. S. (1996). Training for complex decision making: a test of instruction-based on the recognition/metacognition model. in: *Proceedings of the Symposium on Command and Control Research and Decision Aids*.

Fowlkes, J. , Dwyer, D.J., Oser, R.L., & Salas, E. (1998). Event-based approach to training (EBAT). *The International Journal of Aviation Psychology*, 8(3), 209-222.

Helsdingen, A.S. & van den Bosch, K. (1999). Concepts for training tactical command (Report TM-99-A043, in dutch). Soesterberg, The Netherlands.

Klein, G., McCloskey, M., Pliske, R., & Schmitt, J. (1997). Decision skills training. In: *Proceedings of the Human Factors Society 41th Annual Meeting*,(pp. 182-185). Held at: Albuquerque, AL. Santa Monica, CA: The Human Factors Society.

Lussier, J. (2003). Intelligent tutoring systems for command thinking skills. *ARI Newsletter*, volume 13, number 1, 7-9.

Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (1996). The role of shared mental models in developing team situational awareness: implications for training. *Training Research Journal*, *2*, 85-116.

Initial Evaluation of USAF Security Forces Distributed Mission Training (SecForDMT)

Dr. Joseph Lloyd Weeks
War Fighter Training Research Division,
United States Air Force Research Laboratory,
6030 South Kent Street, Mesa, AZ, USA 85212-6060
Tel: +1 480 988 6561 Ext 249 Fax: +1 480 988 6285
E-mail: joseph.weeks@williams.af.mil

ABSTRACT

United States Air Force security forces have a central role in force protection. Their missions include military police services, installation security, air base defense, military working dog functions, and combat arms training and maintenance. Surveys indicate that among the hundreds of tasks in the career field, "Directing security forces" is rated highest on training emphasis (Weeks, Garza, Archuleta, and McDonald, 2001). As a result of established needs and technology opportunities, the Air Force Research Laboratory is conducting research and development of a computer-based simulation capability called Security Forces Distributed Mission Training (Weeks and McDonald, 2002). The capability is designed to support training in decisionmaking, leadership, and team coordination. It allows an instructor to start a simulation exercise on trainee computer workstations connected via a local area network. Simulation software supports the interaction of trainees with each other and with computer-generated forces (CGFs) that imitate behavior of enemy, neutral, and friendly troops and civilians. Radio functions allow multi-channel communication among instructors, trainees, and CGFs. A major design objective is to develop a simulation control interface that instructors and trainees can directly use so the costs of an on-site computer technician can be avoided. The purpose of this paper is to describe outcomes from an evaluation of the usability of the simulation control interface and the validity of computer models.

INTRODUCTION

United States Air Force security forces are responsible for military police services, installation security, airbase defense, military working dog functions, and combat arms training and maintenance. Whether security forces discover an improvised explosive device in a car or confront representatives of a non-governmental organization at an entry control point, situations are reported to a command post. Once situations are reported, quick and accurate decisions by security forces leaders are critical to handling the situation properly and fundamental to the protection of personnel and assets.

Command post exercises are routinely conducted to train personnel to respond to diverse situations. Nevertheless, the command and control of security forces continues to be recognized as a high-emphasis training area (Weeks, Garza, Archuleta, and McDonald, 2001). As a result of established needs and technology obstacles, the Air Force Research Laboratory is conducting research and development of a simulation capability for training leadership, decisionmaking, and team coordination (Weeks and McDonald, 2002). An illustration of the current capability is

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

presented at Figure 1. It consists of standard personal computers connected by a local area network. To reduce acquisition and maintenance costs, visual displays consist of standard computer monitors rather than immersive displays. A trainee station consists of a computer processor, computer monitor, keyboard, mouse, and radio headset and microphone for communications with instructor, other trainees, and computer-generated forces (CGFs). The instructor station consists of one computer for controlling the simulation exercise and recording it for after-action reviews. In Figure 1, trainees are illustrated as a shift leader and subordinate flight leaders but could alternatively be a law enforcement desk sergeant and field officers, a flight leader and subordinate squad leaders, or a defense force commander, operations officer, and field supervisors. The capability is being developed as an expandable, multi-echelon, command and control, training device.

During this initial stage of development, the capability is being tailored to support training for air base defense. During application, an instructor would present on the local area network a terrain model for the area of operations (AO). Trainees would simultaneously view the AO, collaborate in conducting security vulnerability analyses, and develop a security plan. Based on consideration of threats, vulnerabilities, and available resources, trainees would develop the plan by creating and positioning computer models. The simulation capability provides computer models representing sensors, obstacles, fighting positions, vehicles, communications, and semi-automated, CGFs. Computer-generated forces are designed to move, sense, shoot, and communicate. After the plan is completed, trainees would share it with the instructor. The instructor would evaluate the plan and could task CGFs to present security situations to test the plan. After the instructor tasks CGFs to present security situations, the exercise would begin. A situation report from a friendly CGF would be the initiating action for trainee decisionmaking and team coordination. All communications via radio microphones and actions occurring on the visual display would be recorded for after-action reviews. Communications from CGFs to trainees and among trainees would be the focus of student evaluation.

Research objectives include development of a usable control interface, realistic behaviors for CGFs, development of simulation exercises to support learning objectives, and evaluation of system usability, model validity, and training effectiveness. The research and development project includes multiple field evaluations with participation of end users. The development strategy is to collect feedback and apply it to refinement of the capability in an effort to accelerate transition to the field.

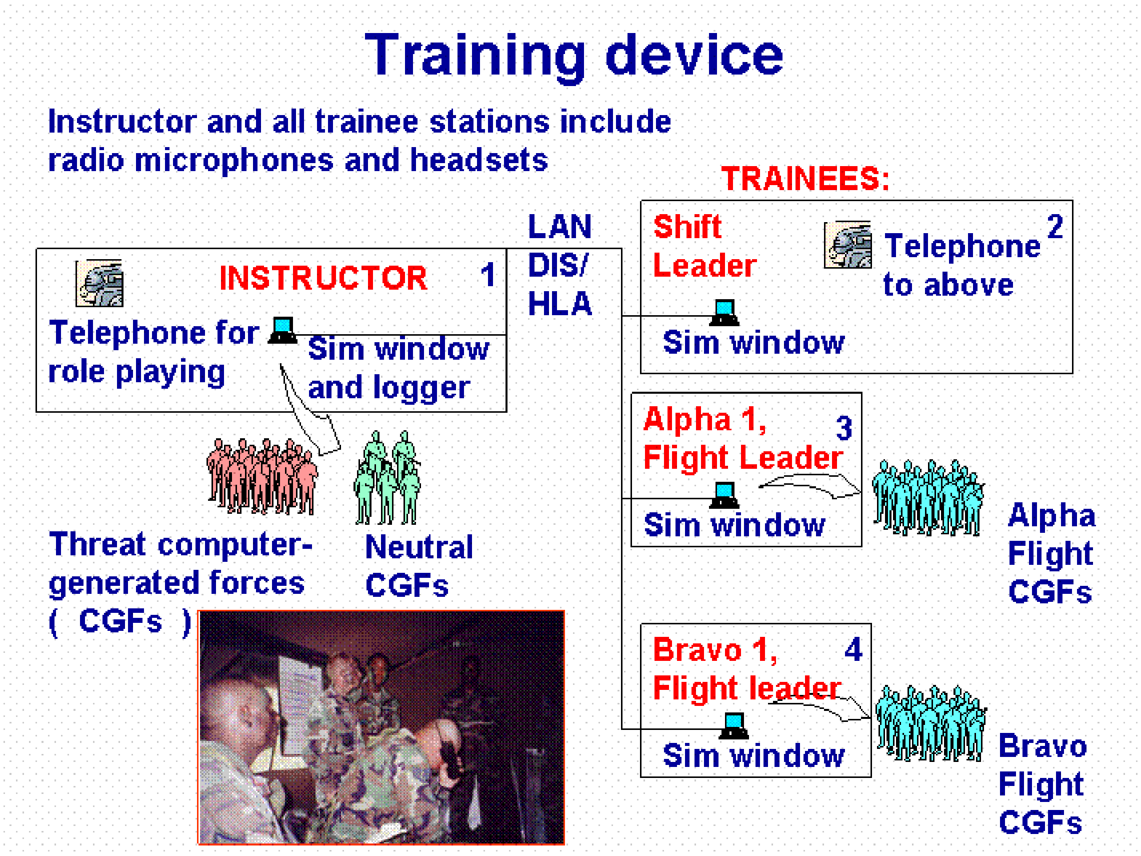


Figure 1. The training device

Usability assessments include measures of time required to train participants to use the capability and instructors' ratings of the usability of the control interface. Instructors' model validity assessments focus on computer models for obstacles, CGFs, and sensors. Evaluations conclude with global ratings of the value of the device for training decisionmaking and team coordination that underlying mission planning and execution of the defense. A preliminary evaluation was conducted and found to be too long in duration (Weeks and McDonald, 2003). As a consequence, the original evaluation procedure was modified. The purpose of this paper is to describe outcomes from the revised procedure for the usability evaluation and future plans for evaluating training effectiveness.

APPROACH

Brewer, Armstrong, and Steinberg (2002) state "usability testing … verifies that a system or product design meets our expectations for usefulness and satisfaction before we move into production (p 403)." They define usability as "the degree to which the design of a device or system may be used effectively and efficiently by a human (p 403)" and point out that the important issue in arriving at a definition is how to measure usability so that measurements can be used to improve the design. Brewer, et al. (2002) outline three general approaches to usability evaluation including surveys (using self-report data collection methods), usability inspections (specialists scrutinize a design according to a systematic approach and judge its acceptability against certain criteria), and experimental tests (based on quantifying operator performance using controlled data collection techniques).

Usability inspection best describes the approach. The specialists were instructors assigned to different security forces training squadrons. They were first trained to use the simulation control interface; then, they evaluated the interface on selected criteria, identified problems, and recommended improvements. Their evaluation concluded with observation of the performance of computer models and ratings of model validity. These specialists were not experts in human factors engineering, but they did represent a final authority for usability, the end user. Trainees are also end users of the simulation capability. They are expected to control the simulation interface and create computer models for sensors, obstacles, weapons, CGFs, and communications. Trainees participated in the usability evaluation by providing baseline training times for simulation control tasks representative of those they would perform during an exercise. The complete training system consists of several computers linked by a local area network. However, for the usability evaluation only one laptop computer was used. The strategy is to modify the control interface and computer models on the basis of change recommendations before taking the complete training system to the field for evaluation.

To describe the usability inspection approach, Brewer, et al (2002) present an example of a computer interface. It is a single dialogue box. The usability issue is whether to position control buttons on the bottom left or bottom right of the dialogue box. Compared to the dialogue box described by Brewer, et al (2002), the interface evaluated here is huge. It consists of over 450 different controls including menus, tools, dialogue boxes, and intermediate control windows. One recommended approach to usability testing is based on presenting a control interface to end users, not informing them how to use it, observing if they can deduce how to use it, and requesting feedback concerning improvements (Andre, personal communication, 2003). Although this would be a useful approach for the interface described by Brewer, et al. (2002), it was not used here. In addition to the great number of interface controls, most participants had no experience with simulation capabilities like the one evaluated. For them, it was a novel experience; so it was impractical to adopt a discovery learning approach. To minimize the duration of the evaluation period while obtaining meaningful input, it was necessary to familiarize participants with the interface in advance.

The instructor's control interface used for evaluations was delivered with Version 2.0 of the simulation software and is presented at Figure 2. It consisted of a simulation window in which a terrain model is presented; a menu bar consisting of menus, menu options, and menu option labels; a tool bar consisting of tool bar buttons and information windows; a pop-up menu presented in the simulation window and accessed through a right mouse click, a mouse, and a standard computer keyboard. Radio microphone and headset and simulation logger were not included. These interface devices will be tested during evaluation of the complete training system.

One of the greatest obstacles to incorporating such simulations into formal training is support costs. For simulations currently available, on-site technicians are required to design and develop exercises in support of learning objectives and be present during the instructional event to control the exercise, serve as role players, task CGFs, and provide simulation replays to

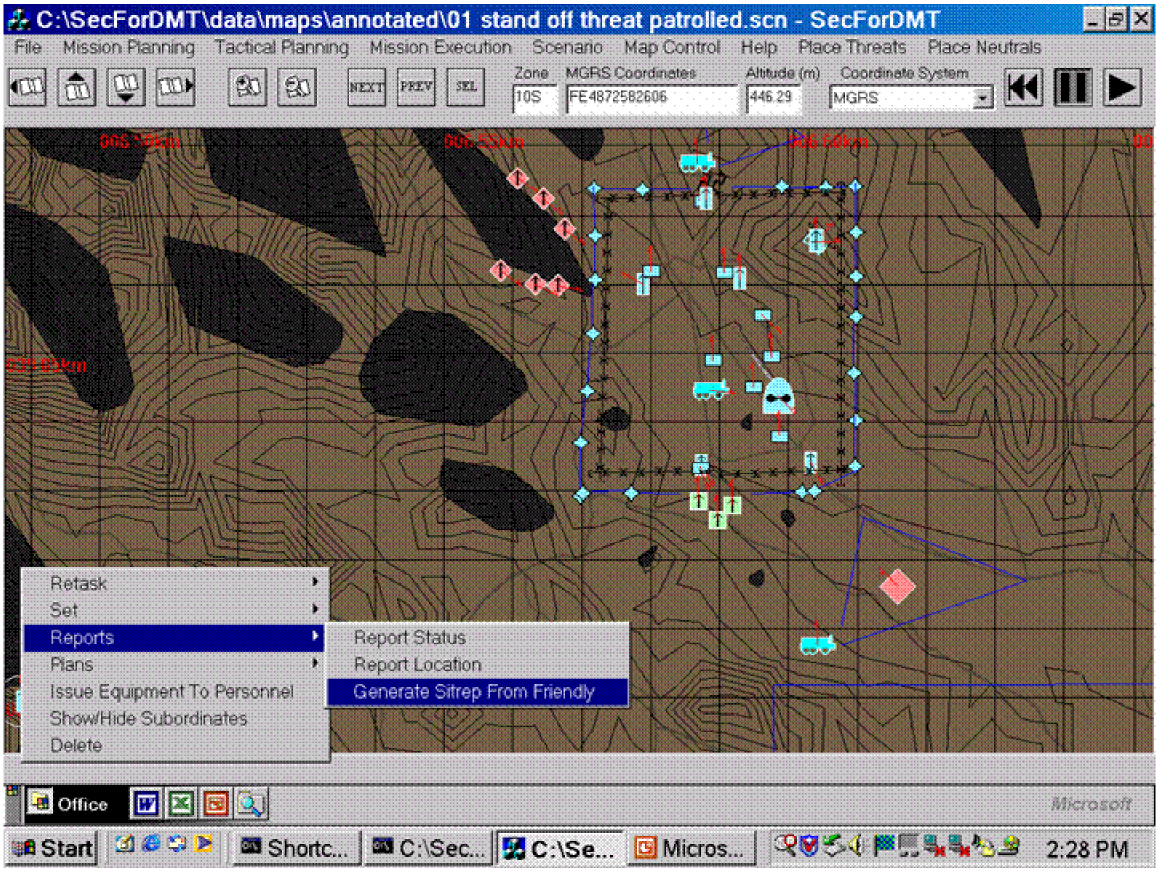


Figure 2. The simulation control interface

support after-action reviews. Support costs are a barrier to fielding simulation technology for formal training; hence, design and development of a simulation control interface that can be directly used by instructors and trainees is a critical technology obstacle. At the beginning of the project, a usability standard was established as the time required to train participants to use the device. The expectation is the shorter the training time; the greater the usability. The intent is to compare the usability standard with observed training times as a guide for system improvement. Training time standards are instructors will be trained to use the device in 2 hours and students will be trained in 30 minutes.

To begin the evaluation, each participant was presented a briefing describing the complete training system and the purpose of the evaluation. After the briefing, they were asked to read and sign a disclosure and consent form and to complete a background questionnaire to obtain information about their time in service, training, and experience with computers. The evaluation was conducted one participant at a time. This approach minimized adverse impact on day-to-day activities of the training group that could have occurred if several participants were tasked to support the evaluation in mass. Each participant was trained to use the control interface and their training time was recorded. It was explained to each participant they were not being evaluated; rather their training time was being measured to estimate training time for the device. Threshold training time was defined as the maximum time that could be allocated for a user to be trained to use the device. After training on all tasks, instructors were asked to estimate maximum training time for instructors and students separately.

The simulation control interface included over 450 control options. Rather than attempting to train participants on all control options, samples of tasks were selected to represent those likely to be performed during a simulation exercise. A task consisted of the use of selected controls to accomplish a purpose. Task selection involved a trade-off between practicality and an exhaustive evaluation. A balance was sought between what participants might regard as an intolerably long evaluation period and evaluation of all control options. The number of tasks was constrained by the length of the evaluation so that the evaluation for each participant did not exceed four hours. Instructors were trained on 18 tasks and students were trained on 13 tasks. Trainee tasks are a subset of instructor tasks and exclude tasks only instructors would perform like creating threat and neutral CGFs. Instructor and trainee tasks are non-random samples from the population of tasks. Although the main purpose was to evaluate usability, the evaluation was the first opportunity for instructors to observe the training device; so a secondary goal was to make the experience meaningful. Rather than randomly sampling and presenting tasks, they were carefully selected and sequenced for meaningfulness. Tasks for both instructors and trainees represent three categories consisting of control of the simulation window, placing resources, and tasking CGFs. If more tasks had been selected, total training time would have been greater. However, the more tasks selected; the greater the length of the evaluation period, and the less likely it would be to obtain participants' willing cooperation and meaningful input. Instructor and trainee tasks and control options used for the evaluation are described in the Appendix.

Measures of training time were obtained separately for each participant. The participant was told how to perform each task, she or he was showed how to do it, and asked to independently perform the task with assistance. They were asked to indicate when they had learned to use controls for the task. When the participant stated she or he had learned to perform the task, training time was declared complete. Immediately after training for each task, the participant was asked to perform the task independently. It was noted whether it was performed with or without assistance. It was assumed that if the participant satisfactorily performed the task without assistance, they had learned to use the controls. If the participant asked for assistance, recorded training time for that task for that participant was doubled. Task training time was cumulated over all tasks to obtain an estimate of training time for each participant. Training time was averaged over trainees and instructors separately to estimate total training time for each group.

Trainees did not evaluate the device. Evaluations were conducted only by instructors. After training for each task, the instructor rated the controls used to perform the task on clarity, effectiveness, efficiency, and simplicity; always in that order. Clarity was defined as the degree to which interface controls were clear and understandable. Effectiveness was defined as the degree to which interface controls allowed the task to be performed. This factor provided an opportunity for instructors to recommend needed functionality. Efficiency was defined as the degree to which the controls used in performing the task allowed quick performance. Simplicity was defined as the degree to which the logic of using the controls was complex or easy to understand. Extremes of the rating scale for simplicity were anchored with verbal anchors, "Extremely High Simplicity" and "Extremely High Complexity". Rating scales for clarity, effectiveness, and efficiency were like the one described in Figure 3 except the applicable factor was inserted. The process of obtaining ratings guided the instructor to think about specific criteria for usability and provided indicators of order relationships among tasks for each factor. After rating the control interface for a task on a factor, instructors were asked to identify problems and recommend improvements.

0	= "Do not know"
1	= "No clarity "
2	= "Extremely Low clarity"
3	= "Very Low clarity"
4	= "Below Average clarity"
5	= "Average clarity"
6	= "Above Average clarity"
7	= "High clarity"
8	= "Very High clarity"
9	= "Extremely High clarity".

Figure 3. Example of one of four usability rating scales

When task training and usability ratings were completed, each instructor evaluated the validity of computer models for obstacles, CGFs, and sensors. They observed five simulation runs that showed the performance of selected computer models and rated the validity of what they observed. The validity rating scale was like the one described in Figure 3 except the word, "Validity", was inserted for the underlined factor. In addition, a sensor expert and a weapons expert reviewed specifications used to develop computer models for sensors and weapons and corrected specifications as appropriate. After observing simulation runs, instructors were asked if they believed the device would support learning objectives for mission planning and execution of the defense and whether or not they believed the simulation capability would add value to training.

RESULTS

A total of 13 security forces instructors participated in the evaluation (3 instructors from the 342nd Training Squadron, 7 instructors from the 343rd Training Squadron, and 3 instructors from the 96th Security Forces Ground Combat Training Squadron). Each instructor dedicated approximately 4 hours to the evaluation including breaks. Participants also included 10 trainees from the 343 Training Squadron who had recently graduated from security forces initial-skills training and were awaiting assignments.

On average, the instructors were 31 years of age. There were 3 captains, 1 senior master sergeant, 1 master sergeant, 4 technical sergeants, and 4 staff sergeants. All enlisted personnel serving as instructors possessed the journeyman skill level or higher. They had an average of 11 years and 6 months of service and 2 years and 2 months in their current position. They indicated they spend an average of 35 hours per week using computers and in the preceding year played computer games an average of 3 times. On average, trainees were 19 years of age. Five trainees possessed the rank of airman first class and 5 possessed the rank of airman basic. All trainees possessed the apprentice skill level. They had an average of 6 months of service. They indicated they spend an average of 4 hours per week using the computer and in the preceding year played computer games an average of 3 times.

Figure 4. Average task training time for 13 instructors
(Summation over 18 tasks = 1 hr 17 min)

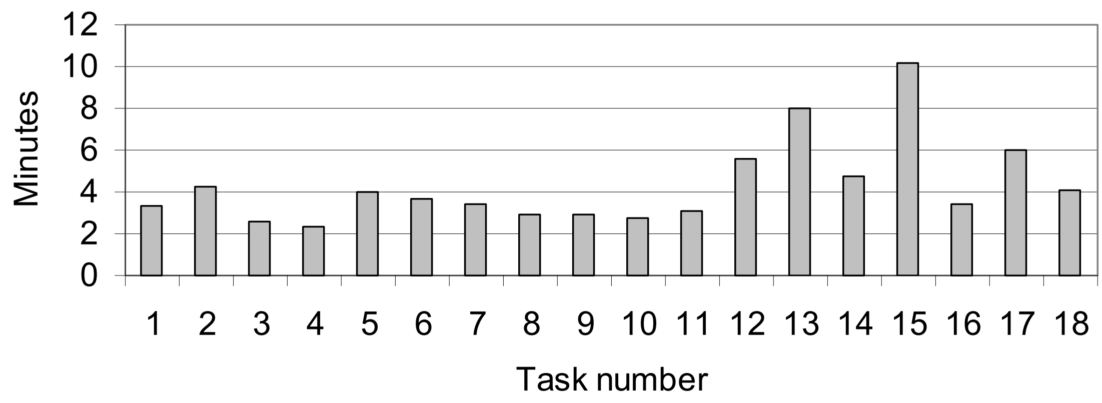


Figure 5. Average task training time for 10 trainees
(Summation over 13 tasks = 56 min 2 s)

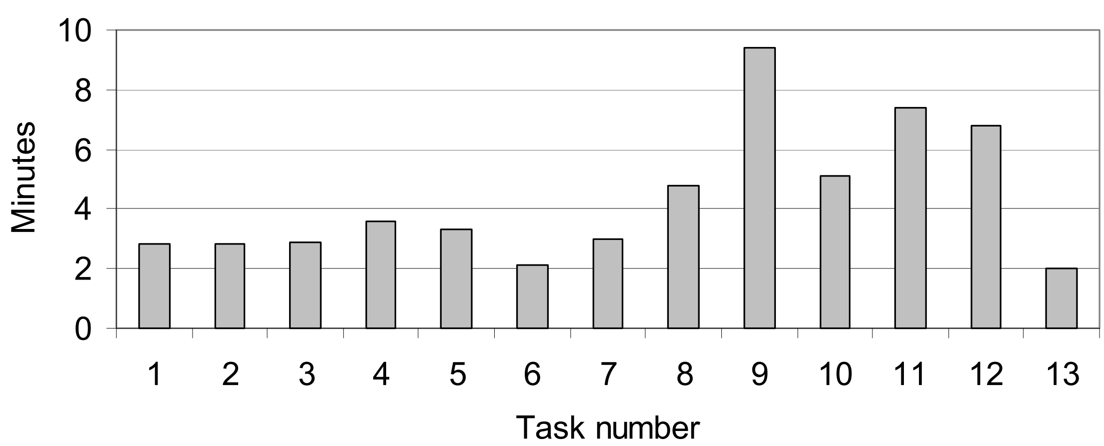
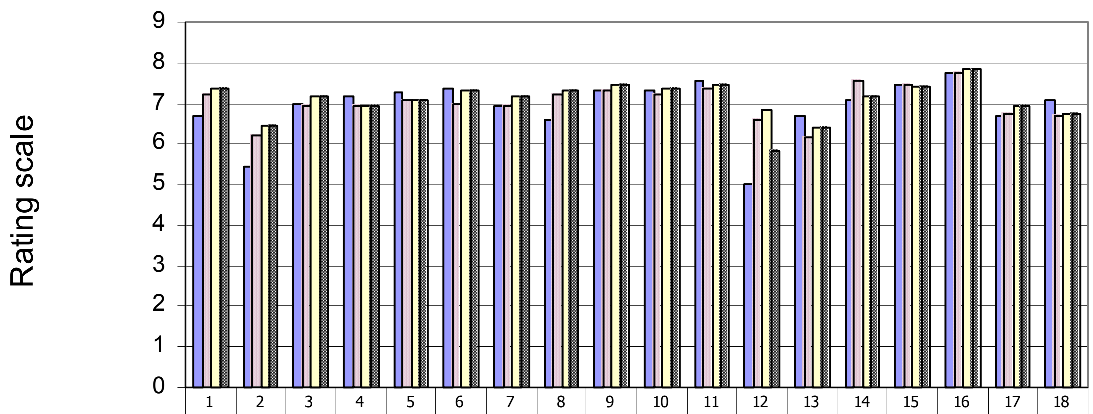


Figure 4 presents the observed task training time for 18 tasks averaged over instructors. Most tasks required 5 minutes or less training time. The instructor training time usability standard was 2 hours. Training time for instructors summed over the 18 task sample is 1 hour and 17 minutes. Average training time for the 18 task sample fell below the pre-established standard. After task training, each instructor was asked to estimate the maximum amount of time that could be allocated for instructors to be trained to use the device. The minimum was 1 hour, the maximum was 8 hours, and the median was 4 hours.

Figure 5 presents the observed training time for 13 tasks averaged over trainees. Most of the tasks required 5 minutes or less training time. For trainees, the training time usability standard is 30 minutes. Training time for students summed over the 13 task sample is 56 minutes. Average student training time for the 13 task sample exceeded the pre-established standard. Each instructor was asked to estimate the maximum amount of time that could be allocated for students to be trained to use the device. The minimum was 1 hour, the maximum was 8 hours, and the median was 4 hours.

For instructors, the procedure was to learn to use interface controls then rate the control interface on clarity, effectiveness, efficiency, and simplicity. Figure 6 presents usability ratings for each factor. Instructors were told that the collection of interface controls used to perform the task were the target for each rating. For each usability factor, a rating value of 5 represents an "Average" rating. Aggregate ratings for all tasks and factors were rated average or above. These

Figure 6. Average interface usability



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Clarity	6.7	5.5	7.0	7.2	7.3	7.4	6.9	6.6	7.3	7.3	7.5	5.0	6.7	7.1	7.5	7.8	6.7	7.1
Effectiveness	7.2	6.2	6.9	6.9	7.1	7.0	6.9	7.2	7.3	7.2	7.4	6.6	6.2	7.5	7.5	7.8	6.8	6.7
Efficiency	7.4	6.5	7.2	6.9	7.1	7.3	7.2	7.3	7.5	7.4	7.5	6.8	6.4	7.2	7.4	7.8	6.9	6.8
Simplicity	7.4	6.5	7.2	6.9	7.1	7.3	7.2	7.3	7.5	7.4	7.5	5.8	6.4	7.2	7.4	7.8	6.9	6.8

Task

results indicate instructors believed the control interface was easy to understand and effective. It allowed them to accomplish tasks quickly and the logic of operations required for using the controls was easy to understand. Ratings for Task 16 deserve special attention. This task presented the capability for instructors to send situation reports from friendly CGFs to serve as a stimulus for trainee decisionmaking and team coordination. Control options allowed the instructor to open a dialogue box with an editable window in which a situation report could be typed/stored and immediately sent from a selected friendly CGF. The process of sending the situation report revealed the sound dimension of the simulation capability, highlighted audible communications among trainees and CGFs, and illustrated support for training decisionmaking and team coordination. Task 16 was rated highest on usability. Even though instructors provided favorable usability ratings for the interface, they identified several problems with the interface and made important recommendations for improvement.

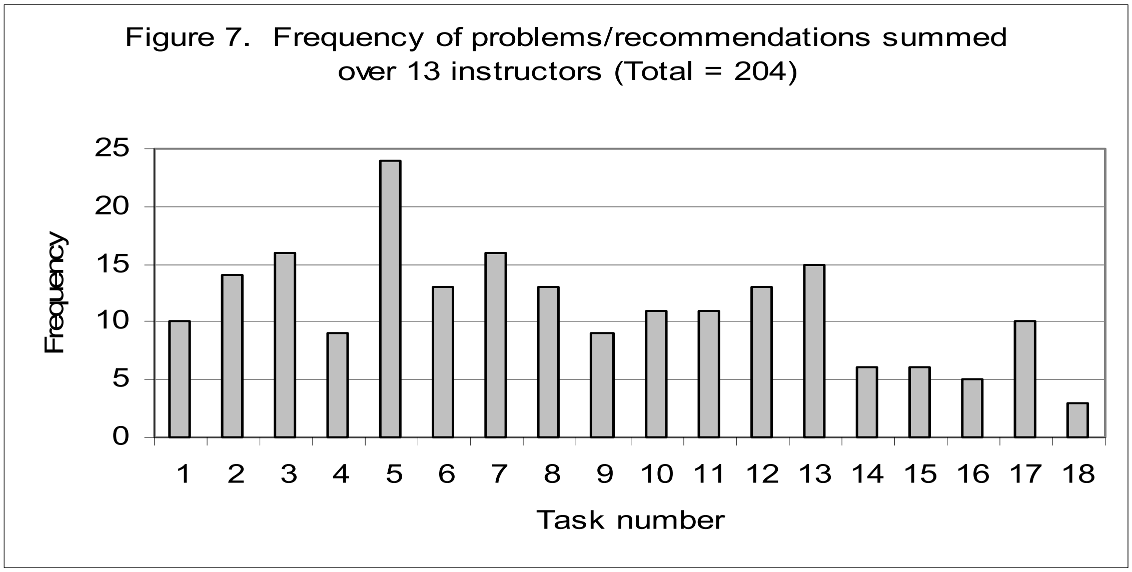


Figure 7 presents the frequency of interface problems/recommendations summed over instructors for each task. They provided a total of 204 change recommendations. Change recommendations were immediately consolidated, prioritized, and shared with development engineers for use as a guide for system development. Examples of change recommendations are presented below:

- Move as many options as possible from tool bar menus to right-click menus.
- Provide a reference point for determining the level of zoom.
- Make it easier to select the end of a range and bearing line.
- Reduce the number of steps required to display terrain contours.
- Reduce the number of steps required to create sensors, CGFs, etc.
- Collapse aircraft menu options into aircraft priority level options.
- Provide an information window to allow instructors to quickly determine the number of weapons, vehicles, radios, and sensors used in the security plan.
- Provide short-cut index for instructor to review all situation reports at once without having to push buttons to display each stored situation report.

After training and usability ratings, instructors observed simulation runs that presented the performance of computer models for obstacles, CGFs, and sensors. Simulation runs included (1) Obstacle delay: A lead, dismounted CGF was tasked to move to a waypoint beyond a single strand of concertina wire and two CGFs were tasked to follow; (2) Obstacle avoidance: A vehicle was positioned in front of a fence and was tasked to move to a waypoint beyond the fence; (3) Fire at will: A dismounted, friendly CGF armed with a M16 was tasked to wait indefinitely with rules of engagement set to "Fire at will'. Rules of engagement for a dismounted, threat CGF armed with an AK47 were set to "Fire at will" and it was tasked to move to a waypoint immediately behind the friendly CGF; (4) Fire if fired upon: A dismounted, friendly CGF armed with a M16 was tasked to wait indefinitely with rules of engagement set to "Fire if fired upon". Rules of engagement for a dismounted, threat CGF armed with an AK47 were set to "Fire at will" and it was tasked to move to a waypoint immediately behind the friendly CGF; (5) Sensors: A dismounted CGF moved on a route past a sensor beam for an active infrared sensor, a trip wire, a passive infrared sensor, a seismic sensor, and a magnetic sensor. In addition, a truck and tank moved on separate routes that led them within 50 meters, 25 meters, and 1 meter of seismic and magnetic sensors.

Figure 8. Computer Model Validity

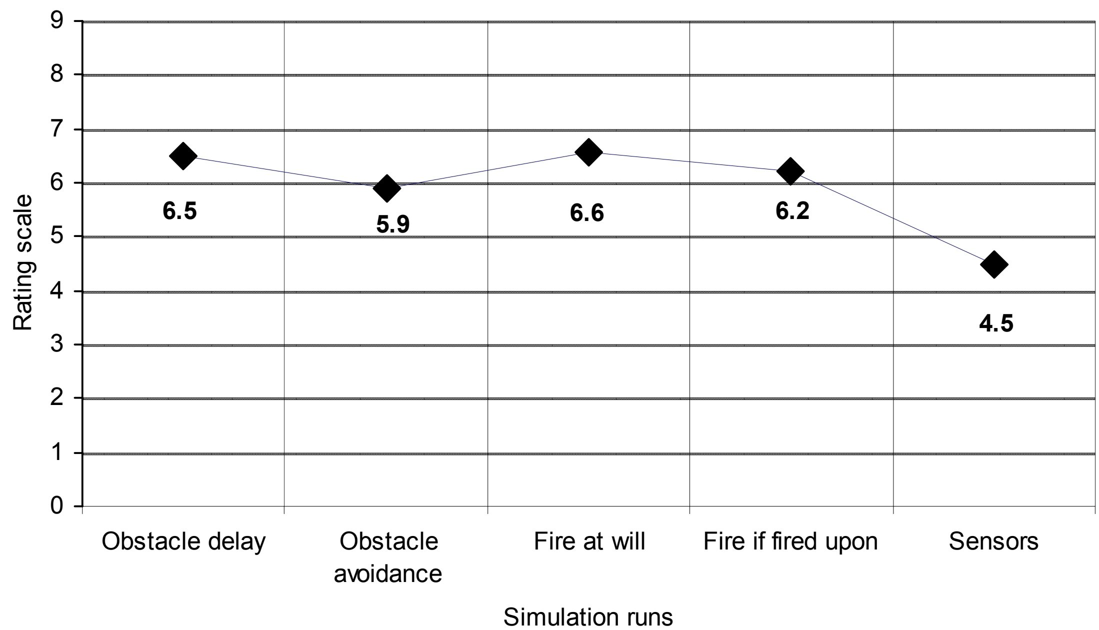


Figure 8 presents average computer model validity ratings for each simulation run. A rating value of 5 represents an "Average validity" rating. Validity ratings for all simulation runs were above average except for sensors. The performance of sensor models was questioned because seismic and magnetic sensors did not detect the presence of vehicles and did not detect the presence of a dismounted CGF at expected distances.

Although there were no weapons validity ratings obtained, the "Fire at will" and "Fire if fired upon" simulation runs did showcase weapons performance. Several instructors commented on weapons effects. For the "Fire at will" simulation run, a friendly CGF armed with a M16 killed the threat CGF armed with an AK47 at approximately 500m. More than one instructor stated that a kill at 500 meters by an average shooter was unrealistic. Under the "Fire if fired upon" condition, a friendly CGF armed with a M16, standing in the open, not in defilade, was hit by a threat CGF armed with an AK47 at 100 meters. The friendly CGF returned fire and killed the threat CGF. The weapons expert stated the effect of the AK47 on the friendly CGF was unrealistically minor and that, in reality, the AK47 would have a more damaging effect. Discussions with software engineers revealed the probability of kill for friendly CGFs was based on the assumption they would be wearing protective Kevlar vests. Revised kill probabilities were obtained and passed to development engineers for system revisions.

After task training and the process of providing usability and validity ratings, each instructor was asked to provide global evaluations. All 13 instructors agreed the device would add value to training and would support learning objectives for security planning and execution of the defense. The officer responsible for managing the security forces officers' command course expressed the desire to have the simulation capability available for the course by the next class.

DISCUSSION

Although training times provide useful baseline estimates of usability, they have limitations. The evaluation did not include the simulation logger or the radio microphone or headset. If tasks for these devices had been included, training time would have been greater. Training time for these devices will be estimated when the complete training system is evaluated. Because it was necessary for reasons of practicality to limit the number of tasks evaluated, training times are underestimates. Only a representative subset of 18 tasks was evaluated for instructors and only 13 tasks were evaluated for trainees. Observed training times are underestimates of total training time for all 450 interface control options. However, knowledge of all control options is not necessary for practical use. How many users of Microsoft Word know how to apply all control options? For instructors, total observed instructor training time for the 18-task sample is an overestimate. During the training period for each task, instructors often discussed problems and recommended needed capabilities. The result was that total observed training time includes time for training and non-training interactions. It would be desirable to obtain a more accurate estimate of training time that excludes time for non-training interactions, includes tasks for use of the radio headset, and simulation logger, and a greater number of representative tasks while maintaining a reasonable time for the evaluation period. Multiple evaluation periods may be necessary if acceptable to participants.

Apart from limitations in observed training time, it is important to understand the potential represented by this capability. If observed training time for instructors were doubled, it would still be less than the training time required for other simulation capabilities available today. Consider the Joint Combined and Tactical Simulation (JCATS). On a separate occasion, the author observed security forces personnel being trained on JCATS to support installation security evaluations. Observation of "user training" for JCATS indicated 7 security forces personnel varying in rank from staff sergeant to master sergeant satisfactorily learned to task CGFs to move, dismounted and mounted, and shoot, direct and indirect-fire weapons, in 1 and ½ eight-hour, training days. JCATs controllers indicated they had learned to operate JCATS over a period of months. If the training time for JCATS is the comparison point, SecForDMT offers a significant advantage by avoiding lengthy training times for users.

Computer model validity assessments indicated instructors believed computer models for obstacles and CGFs were above average in validity. The simulation run that illustrated obstacle performance showed dismounted CGFs being automatically delayed while moving past a strand of concertina wire and a vehicle automatically avoiding a fence strand. The simulation run that illustrated CGF behavior showed a friendly dismounted CGF complying with rules of engagement for "Fire at will" and "Fire when fired upon". The aggregate rating for sensors was below average. This reflects the fact that seismic sensors did not alarm at the expected detection distance and magnetic sensors did not alarm for vehicles.

The usability evaluation provided valuable information for spiral-development; but, no information describing training effectiveness. Estimates of training effectiveness would be required to quantify benefits for conducting cost-benefits analyses which, in turn, would inform acquisition decisions. Estimates of training effectiveness could be obtained by conducting an experiment using controlled data-collection techniques. An initial objective would be to determine if decisionmaking performance at the individual level improves over successive simulation exercises. To reduce resource requirements, a repeated-measures, pre-test/post-test, experimental design would be used to attain the greatest statistical power for the minimum number of participants. At least, 20 individuals would be required to participate. The value of the training device can not be separated from the instructor who applies it and the simulation training syllabus. So, it would be necessary to enlist instructors' participation to develop simulation exercises, define performance standards, deliver exercises, and evaluate trainee performance. Resource requirements would be significant. Procedures for conducting such an experiment are currently being formulated and discussed with representatives of training organizations.

CONCLUSIONS

The capability holds promise for being a highly usable training device. Observed training time satisfied the standard for instructors but failed to satisfy the standard for trainees. Although trainees' observed time of 56 minutes exceeded the goal of 30 minutes, the maximum time judged available for training students to use the device (median of 4 hours) suggests the capability could be assimilated into training. Instructors rated usability average or above average for all tasks and rated the validity of computer models above average for all computer models except sensors. Instructors unanimously agreed the device would add value to training and that it would support learning objectives for security planning and execution of the defense. Most importantly, the usability evaluation resulted in over 200 change recommendations. While the system is being improved in accordance with these recommendations, a plan is being formulated to take the complete training system to the field for the purpose of identifying relevant learning objectives, developing simulation exercises, and identifying performance standards and metrics. This information will inform design of an experiment to quantify training effectiveness.

REFERENCES

Armstrong, S.D. and Steinberg, R.K. (2002) *Usability Testing.* In Charlton, S.G., O'Brien, T.G. (Eds.) H*andbook of Human Factors Testing and Evaluation.* Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, Chapter 18, p 403-432.

Weeks, J. L., Garza, J. S., Archuleta, M. A. and McDonald, L. B. (2001) USAF Security Forces Training Needs. *Proceedings of the 2001 Interservice / Industry Training, Simulation and Education Conference*, Orlando Florida.

Weeks, J. L. and McDonald, L. B. (2002). USAF Security Forces Distributed Mission Training: Evolution of the design concept. In, *Proceedings of Spring 2002 Simulation Interoperability Workhop*, Orlando, FL.

Weeks, J. L. and McDonald, L. B. (2003). Preliminary evaluation of the control interface for a training simulation system. *Proceedings fo the North Atlantic Treaty Organization, Research and Technology Agency, Systems Concepts and Integration Panel, Symposium on Critical Design Issues for the Human-Machine Interface,* Prague, Czech Republic, 19-21 May, 2003.

APPENDIX

INSTRUCTOR TASK LIST (K_i=18)

Task module: Control simulation window

Task 1. Zoom in 4 levels, zoom out 2 levels, pan right, and use Sel/Next/Prev buttons to center map.
Task 2. Determine distance represented by side of a map grid square by interpreting map coordinates.
Task 3. Create, select, move, and orient a range and bearing line and determine distance represented by side of map grid square and delete line.
Task 4. Turn on contour line display and determine elevation with cursor and Altitude window.

Task module: Resource placements

Task 5. Create, select, move, and delete a facility and an aircraft.
Task 6. Create, select, move, and delete an obstacle.
Task 7. Create, select, move, orient, and delete a sensor.
Task 8. Create, select, move, orient, and delete a friendly computer-generated force (CGF) armed with an M9 and a friendly CGF armed with an M2.
Task 9. Create, select, move, orient, and delete a threat CGF armed with an AK47 and a threat vehicle (BMP-2 AFV).
Task 10. Create, select, move, orient, and delete an armed neutral CGF and an unarmed neutral CGF.
Task 11. Create, select, move, orient, and delete a defensive fighting position.
Task 12. Place friendly CGF in defensive fighting position, turn on primary field of fire and effective range and set primary field of fire.
Task 13. Create a terrain box, create an observation point, determine intervisibility from the observation point, delete the terrain box and observation point.
Task 14. Withdraw and assign telephones to facilities and set loop designations.
Task 15. Withdraw and assign radios to entities, set call signs, and receive and transmit frequencies.
Task 16. Create, save, and send a situation report.

Task Module C: Control Forces

Task 17. Create a route and write a task plan for a HMMWV to patrol a route at maximum speed.
Task 18. Retask.

TRAINEE TASK LIST (K_t=13)

Task module: Control simulation window
Task 1. Use Sel/Next/Prev buttons to center map on entities, zoom in 4 levels, zoom out 2 levels, and pan.
Task 2. Create, select, move, and orient a range and bearing line and determine distance represented by side of map grid square and delete line.

Task 3. Turn on contour line display and determine elevation with cursor and Altitude window.

Task module: Resource placements

Task 4. Create, select, move, and delete an obstacle.
Task 5. Create, select, move, orient, and delete a sensor.
Task 6. Create, select, move, orient, and delete a friendly computer-generated force (CGF).
Task 7. Create, select, move, orient, and delete a primary fighting position.
Task 8. Place friendly CGF in primary fighting position, turn on primary field of fire and effective range, and set primary field of fire.
Task 9. Create a terrain box over the primary fighting position, create an observation point, determine intervisibility from the observation point, delete he terrain box and observation point.
Task 10. Withdraw and assign field telephones and set loop designations.
Task 11. Withdraw/assign radios, set call signs, and channels, enable audible radio communications, create reference point, and send a location report.

Task Module: Control Forces

Task 12. Create a route and write a task plan for a HMMWV to patrol a route at maximum speed.
Task 13. Use the Retask menu option to retask an entity to go to a particular point on the terrain map.

Usability Analysis of a Personal Digital Assistant Based Data Collection Tool for the Shipboard Training Environment

Dr. Robert C. Allen
NAVAIR Orlando Training Systems Division
12350 Research Parkway
Orlando, FL 32826-3275
USA

Mr. Paul J. Hession
Sonalysts, Inc.
12501 Research Parkway
Orlando, FL 32826-3224
USA

Ms. Eleni D. Kring
Dynamics Research Corporation
3505 Lake Lynda Dr.
Suite 100
Orlando, FL 32817-
USA

Summary

Researchers for the U.S. Navy have developed multiple instructor aides for performance measurement hosted on hand-held computers such as pen tablet computers. This technology provides a potential solution to the challenge of supporting training in complex, data intensive shipboard environments. However, Hand-held computers are relatively expensive and can be cumbersome in the confined spaces found in such environments. Therefore, the U. S. Navy is investigating the use of more portable, lightweight data collection tools such as Personal Digital Assistants (PDAs). Hardware and software limitations associated with these devices exist, including limited screen real estate and memory. The challenge for the Naval training and human factors communities is to develop training applications for PDAs that are relevant to shipboard users and that also apply sound human factors and usability principles. The primary purpose of this paper is to describe a usability analysis of a training application loaded onto a Pocket PC. The analysis included heuristic evaluations, user testing sessions, and redesign recommendations. The target audience of the application is U.S. Navy shipboard instructors, who would use the application to prebrief a training audience, to collect data during an exercise, and to debrief the training audience.

Background

In order to fully understand the usability evaluation data reported in this paper, it is necessary to first describe to the reader the forces that led to the development of the subject of this paper, that is, the Personal Digital Assistant (PDA) training application. These forces include the changing training environment in the U.S. Navy, the Navy's implementation of a training system and methodology (Battle Force Tactical Training or BFTT and Objective Based Training or OBT, respectively), the development of software that supports this training methodology (the Afloat Training Exercise and Management System or ATEAMS), and the development of hand-held computer devices that support shipboard data collection. Note that an acronym list is provided at the end of this document to help the reader more easily understand the terms contained herein.

The U.S. military is attempting to reduce both cost and manning while concurrently maintaining operational readiness. In order to meet these conflicting goals, the U.S. Navy is investigating various strategies to augment or supplement training. For example, one approach uses embedded training systems to both simulate a theater of war and stimulate trainees' instruments to display operationally realistic data. Such systems are capable of capturing data that was not possible to capture through previous data collection methods (e.g., paper and pencil). However, the amount of data that can be displayed/given to an instructor, either real-time or during an After Action Review, can be overwhelming. To help aid the instructor capture

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held in Genoa, Italy, 13 – 15 October 2003, and published in RTO-MP-HFM-101.

training-relevant data in highly complex and data-rich training environments, the U.S. Navy has been developing automated data collection tools. For example, automated performance measurement is being implemented under the Navy's Battle Force Tactical Training (BFTT) system. BFTT immerses trainees in a controlled, interactive Synthetic Theater of War environment. The BFTT system is designed to allow trainees to train as they fight, using their operational equipment during a training exercise. The trainees can be onboard ship or in a schoolhouse (see Figure 1) (RCI, 2000).

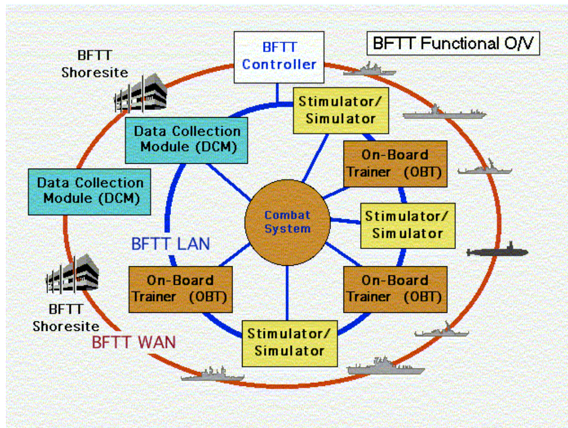


Figure 1. BFTT functional overview. From Battle Force Tactile Training (BFTT) System Overview CD-ROM.

The BFFT effort has defined a 'train by objective' strategy, to be used to improve training onboard ship. This strategy requires the identification of quantifiable training objectives that can be used to accurately rate performance (Lyons & Allen, 2000; Stretton, 2001). Therefore, Commander in Chief, U.S. Pacific Fleet approved a training concept labeled Objective Based Training (OBT). OBT defines the tasks that must be performed, either at the individual, team, or ship level, how these tasks must be performed, and the standards that must be achieved (Lyons & Allen, 2000).

Objective Based Training

The OBT strategy employs Terminal Objectives (TOs), Enabling Objectives (EOs), and Measures of Performance (MOPs) of various warfare areas, such as Combat Systems, Engineering, and Damage Control. Overarching the TOs, EOs, and MOPs are Training Events (TEs). A Training Event is a very high level description of an event that will occur during a training exercise - for example, Employ Firepower - and is composed of one or more TOs. TOs are objectives to which the ship, team or crewmember train. They are high-level objectives that, when achieved, indicate satisfactory accomplishment of the task (e.g., accurately classifying aircraft). EOs are lower level tasks or actions that, when performed correctly, allow the ship, team or crewmember to meet the terminal objective (e.g., was the IFF utilized to classify aircraft?). Multiple EOs may be needed to meet a given TO. MOPs are still lower level tasks or actions that the ship, team or crewmembers need to execute, which then determine whether an EO was performed correctly (e.g., was IFF utilized and all modes challenged to identify air contacts?). Multiple MOPs may need to be executed to meet the EO (see Figure 2; Lyons & Allen, 2000; Stretton, 2001).

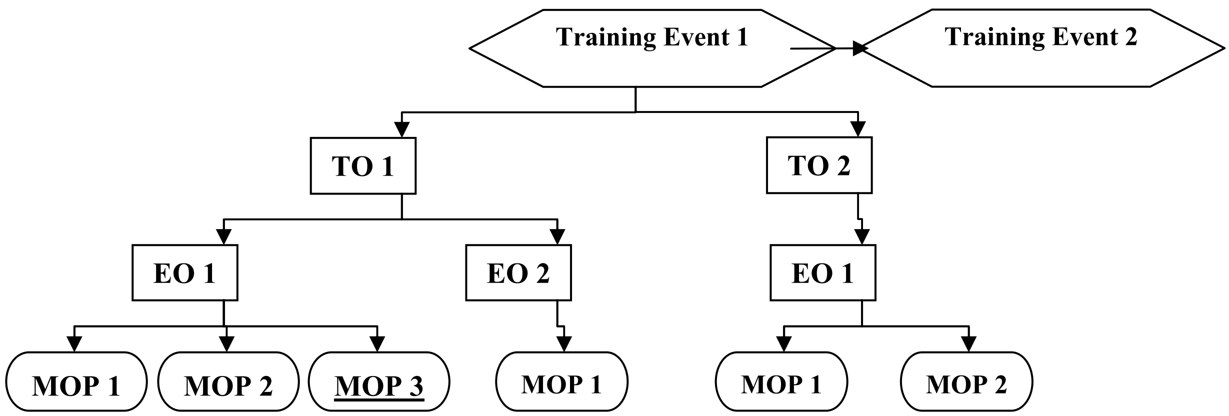


Figure 2. Example of hierarchy of OBT process. Note that TEs are not necessarily sequential.

Afloat Training Exercise and Management System (ATEAMS)

The OBT process was originally implemented manually (paper-based), but this method proved to be labor intensive. Therefore, a requirement to develop an automated tool dubbed the Afloat Training Exercise and Management System (ATEAMS) tool was issued. ATEAMS is a PC-based software application created with Embedded Visual Basic© (eVB). It is designed to manage data relating to basic training of teams and individual Naval crewmembers in both live and simulation-based training exercises, while adhering to the OBT process (Lyons & Allen, 2000). Stretton (2001) notes that ATEAMS

> "provides the capability to conduct training based on pre-defined objectives that are both measurable and traceable. Commands can use several paths for selecting objectives that include: Universal Naval Task List, Fleet Exercise Publications, mission selection, training team selection, watchstation selection, watchteam selection, and querying individual operator performance from previous exercises. This selection process … provides a simplified means to develop training scenarios that are traceable to selected objectives, as well as providing standardized methods to measure team and individual performance. ATEAMS capabilities include:
>
> • Support shipboard training teams
> • Plan training events
> • Generate objective-based training scenarios
> • Identify and generate data collection requirements
> • *Provide the means to gather performance data*
> • *Retrieve and integrate collected data and support debrief*
> • Provide the mechanism to conduct trend analyses
> • Provide feedback to chain of command
> • Provide feedback to schoolhouses
> • Provide feedback to Systems Commands
> • Support the administration of data for the above functions" (emphasis added, p. 1430).

The steps of the OBT process that ATEAMS supports are as follows. First, ATEAMS provides a historical database of trainee and team performance. This data can be used by a Naval command to help determine the training requirements for an upcoming training exercise. It can also assist the Naval command in determining the training audience for a given training exercise, for example the Combat Systems Training Team and/or the Damage Control Training Team. Once the training audience is identified, the command can then select training objectives from the Universal Naval Task List and/or Fleet Exercise Publications databases stored on ATEAMS or they can create their own training objects (Stretton, 2001). ATEAMS can then be used to create OBT Events (TEs) that are embedded within a training scenario package. The package would include elements such as TEs, TOs, EOs, MOPs, exercise location, environmental factors, geopolitical factors, and opposition force composition. The current intention is to then take the data generated by ATEAMS and feed it into the BFTT simulation system, which would use the information to simulate the exercise and stimulate trainees' shipboard instruments. ATEAMS would also provide a means by which instructors can collect data during a training exercise. Once the data is collected, the instructor can use it for debriefing purposes and/or later analysis. The data can also be used to update the individual and/or team's performance history. This history will help commands identify teams' strengths and weaknesses, providing them with a means of more precisely identifying training requirements (see the OBT cycle in Figure 3).

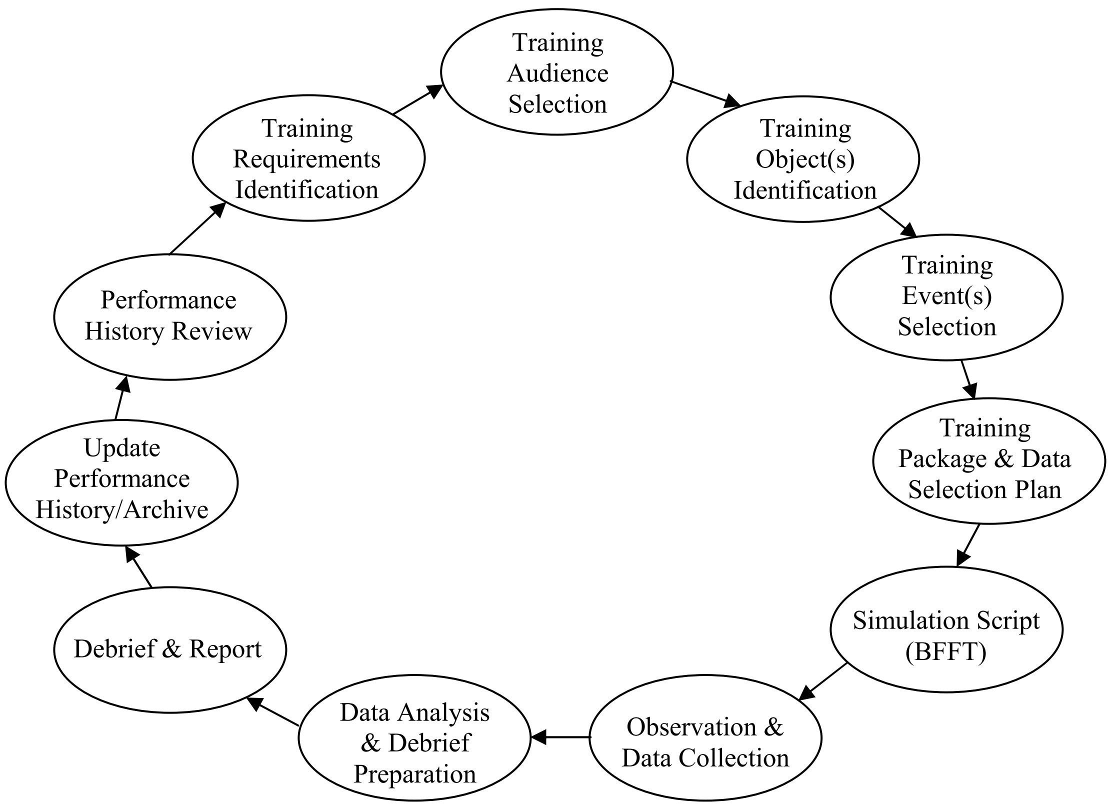


Figure 3. OBT cycle. BFTT drives the simulation system. Adapted from Stretton (2001).

As mentioned, paper-based data collection methods are extremely labor-intensive, both during and after training exercises. In addition, paper-recording methods can lead to inefficient scoring and errors. In the former case, the instructor must take time to flip through sheets of paper during a training exercise, time that could be used to observe performance or provide feedback to the trainee. In the latter case, the instructor may misplace one or more scoring sheets, leading to incorrect data analysis. Therefore, one requirement associated with ATEAMS was the development of a hand-held automated performance assessment tool (Lyons & Allen, 2000). The concept was to link the ATEAMS databases to a hand-held device and download instructor-relevant data (e.g., TOs, EOs, and MOPS) onto this device. An application loaded on the hand-held device would provide an interface that would allow the shipboard instructor to easily capture performance, and other, data. Once captured, this data would be uploaded back into the ATEAMS database to be used for team debriefs, trend analysis, as well as updating ATEAMS' performance history database.

Hand-held Devices

There are important factors that must be taken into consideration when selecting a hand-held device, especially when applied to the military training environment. These factors include memory capacity, the operating system, screen size and type, battery type and storage capacity, ruggedness, and data import and export capabilities (i.e., docking station, cable, modem, and/or infrared connection) (Weber & Roberts, 2000). Initially, pen tablet computers were considered as the platform of choice for the ATEAMS data collection tool. However, pen tablet computers are relatively expensive, are heavy, weighing up to four pounds, and can be cumbersome in the confined spaces found in shipboard environments. Therefore, the Navy is currently investigating the use of more portable, lightweight electronic tools for data collection. One such tool is the Personal Digital Assistant (PDA).

PDA's fall into two general categories: Palm-style organizers and Pocket PCs such as Compaq's iPAQ™ (Consumer Reports, 2001). Pocket PC's are more like mini-computers than traditional PDAs. For example, Pocket PCs have much faster processors, are loaded with familiar applications like Excel™, and, unlike many Palm-style organizers, have color displays.

Within each category, key aspects differentiate one hand-held device from another. Some of these factors include processor speed, memory capacity, battery convenience, display quality, and ease of use. Depending on the model, memory can range from 2 to 32 MB. The battery should be examined for two factors: the life of the battery before replacement or recharging is necessary and the method by which replacement is made. That is, some batteries are rechargeable while others are replaceable. If rechargeable, some PDA manufactures require that the unit be returned to the factory for battery replacement. The quality of the display also differentiates PDAs. Factors in this area include display size, resolution, and color capability.

PDAs and Pocket PCs now have the capability to incorporate multiple tools and functionalities into one relatively lightweight and affordable unit. Many of today's smaller hand-held devices, such as cell phones, have capabilities that include Internet access, digital camera and video/audio recording, ebook, MP3 recording, in addition to cell phone capabilities. Many have the capability to easily record, import, export, and manipulate data. These capabilities are being incorporated into hybrid units (i.e., cell phones with PDA functionalities). Although still requiring usability improvements, hand-held devices can be used for flexible and non-intrusive data collection. Data can be entered through an on-screen electronic keyboard, through handwriting (either natural or script) and/or by voice recording. For recording longer responses or detailed observations, Gravlee (2002) recommends that an external keyboard be used. Hand-held devices can also be loaded with custom data collection software, greatly expanding their utility in training and data collection settings. However, the more functionality added, the higher the associated costs will be to memory, battery life, weight, affordability, and perhaps durability and usability. Nonetheless, the number of hand-held devices with high functionality as well as affordability is growing.

Compared to pen tablet computers, PDAs are lightweight and inexpensive. However, the pen tablet computer has higher computational power, storage capacity, and a larger screen. If the PDA is the platform of choice for the Naval training community, then the challenge for the training application developer is to develop PDA applications that can compensate for the devices' weaknesses and/or capitalize on its strengths. One way to help ensure this is through the use of sound human factors and usability analyses during the development of the training application. For example, a usability analysis can determine whether a PDA-based training application is easy to use or if the displayed graphics and text are readable. Ease of use and readability are important; else the instructor, while attempting to step through a complex application or decipher graphical/textual data, may miss critical events occurring during a training exercise.

To test the feasibility of using a PDA for shipboard training exercises, researchers at Sonalysts, Inc., in conjunction with NAVAIR Orlando Training Systems Division, developed a prototype data collection application ATEAMS PDA (APDA). Through this application ATEAMS, hosted on a PC, could be synched with a PDA. Once synched, training-relevant data such as TOs, EOs, MOPs, and Rules of Engagement (ROE), could be downloaded from ATEAMS to the PDA.

For shipboard exercises, the current intention is that each instructor would have a PDA containing data that would be applicable only to the team that the instructor was evaluating (e.g., engineering or the combat information center). The instructor could use the information stored on the PDA for briefing purposes, for referencing scenario-related material during the training exercise, as well as for capturing data through APDAs' Graphical User Interface (GUI). Once collected, the instructor could use the tool to access the captured data for debriefing purposes. The data could then be uploaded back into ATEAMS databases for later analysis and for updating a team's performance history.
Once the ADPA application prototype was completed, researchers at NAVAIR Orlando Training Systems Division subjected it to a usability evaluation. Usability evaluations can reveal critical and non-critical design flaws in hardware and software and, therefore, was chosen as one method for evaluating the APDA application.

Usability

Usability reflects the extent to which users of a given system can use the functionality of that system. Usability has many components, including how easily a system can be learned (learnability), how easy a system is to remember (memorability), the degree of efficiency that can be obtained after the user has learned the system (efficiency), the error rate (errors), and the subjective satisfaction of using the system (satisfaction) (Nielsen 1993). A usability evaluation generally consists of four phases: a task analysis, a heuristic evaluation, user testing session(s), and design or redesign recommendations. The current evaluation falls primarily under the last three phases of the usability process. That is, due to resource limitations, a task analysis was not conducted. However, SMEs were consulted regarding the nature of training onboard ship. This knowledge was applied to the design of the usability testing sessions.

Method

The current evaluation included two heuristic evaluations, user testing sessions, and redesign recommendations. The evaluation was conducted at NAVAIR Orlando Training Systems Division, located in Orlando Florida, in February 2001.

Participants

Three participants took part in the evaluation. All were male and had prior Navy service (two retired Chiefs, one retired Captain) with backgrounds similar to APDA end-users. All are from the surface community with 15 to 25 years of experience in the training community (mean = 20 years). All participants indicated a high or very high level of experience with Windows®-based applications.

Materials and Equipment

Four PDA models were considered for the current evaluation. The iPAQ™ H3650, manufactured by Compaq®, was chosen as the test bed for the ATEAMS PDA software (see Figure 4). This device was

Figure 4. Compaq® iPAQ™ with and without slipcase.

selected because, of four devices considered, it was the only Pocket PC. Pocket PC's are more like mini-computers than traditional PDAs. For example, Pocket PCs have much faster processors, are loaded with familiar applications like Excel™, and, unlike many Palm-style organizers, have color displays. The iPAQ™ H3650 runs on Microsoft Windows® CE with an ARM SA1110 processor with 31.15 MB of main memory.

APDA Application Windows. There are three main windows associated with the APDA application – these are labeled Prebrief, Collecting Data, and Debriefing. Four additional windows can be activated through the Collecting Data window. These include a Timeline window, an Assessment window, a window labeled Incomplete?, and a Comment window. A brief description of the functionalities associated with each window follows.

Prebrief Window. This window contains information that an instructor may want to access before or during a training exercise. This includes Training Team Assignment (e.g., Combat Systems Training Team), Trainees (names & rates), Timelines (displays the onset time of each TE), Lessons Learned (from previous training exercises), Safety, Rules of Engagement (ROE), Scenario Summary (displays the Training Package name, Mission Statements, Current Situation, and Tactical Objectives), and Objectives (displays the TEs, TOs, EOs, and MOPs, formatted in a tree view control) (see Figure 5).

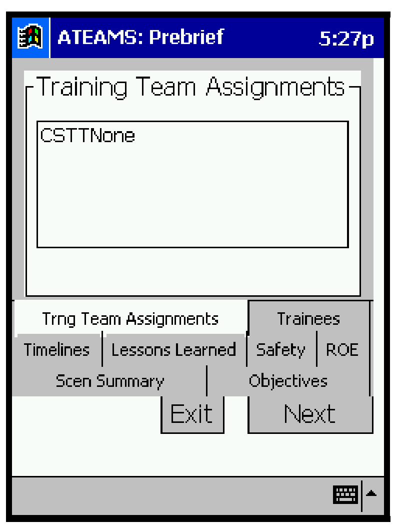


Figure 5. Prebrief window.

Collecting Data Windows. At the top of the Collecting Data window are two radio buttons labeled Timeline and Assessment, which allow the user to toggle between a window displaying the timeline of TEs and a data collection window. The Timeline window contains a list of the start time of all TEs, given in scenario time, using an hours:minutes:seconds format. The instructor can use the timeline information to better manage his time, allowing him to focus on the trainee(s) whose performance will be affected when the start time of a given TE is reached. The Assessment window is used to collect performance data during a scenario run. The Assessment window displays the current TO, EO and associated MOPs. The TO or EO can be viewed by clicking a gray-colored toggle command button (currently set at TO as seen on the middle window in Figure 6).

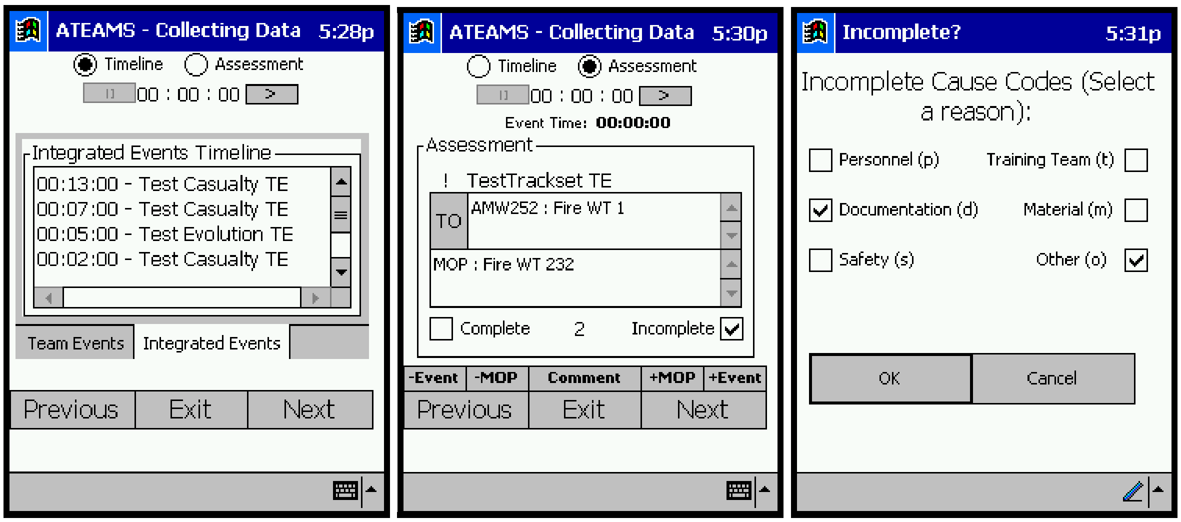


Figure 6. Left to right – Timeline, Assessment and Incomplete Cause Codes windows.

The instructor uses the Assessment window to rate trainee and team performance. To accomplish this, checkboxes labeled Complete or Incomplete are used. If the Complete box is checked, the instructor is indicating that an MOP, relating to a given EO and TO, has been completed. If the Incomplete box is checked, a window labeled Incomplete? appears. Through this window, the instructor can check up to six cause code checkboxes to explain why the MOP was not completed (see Figure 6). Therefore, the performance measurement currently provided by ATEAMS is essentially dichotomous – either a trainee did, or did not, complete an MOP task. After checking the appropriate cause code checkbox(es), and clicking the OK button, the instructor is returned to the Assessment window. Command buttons, located near the bottom of this window and labeled with a plus (+) or a minus (-) sign, are used to display the next or previous TO's, EO's and MOPs. Scroll bars can be used to view the text of long TO's, EO's or MOP's. Navigation buttons, labeled Previous, Next, and Exit, are located at the bottom of both Collecting Data windows. These buttons are used to close one main window and open another (e.g., when the Prebrief window is open, clicking Next will close the Prebrief window and open the Collecting Data window).

When the Comment button, located on the Assessment window, is clicked, a Comment window appears. A comment can be entered through the Pocket PC keyboard or Microsoft's® block or letter recognizer software. The latter method requires the instructor to learn how to write alphanumeric characters based on the software's rules. Each character is written, one at a time, in an input panel area (see Figure 7). When text

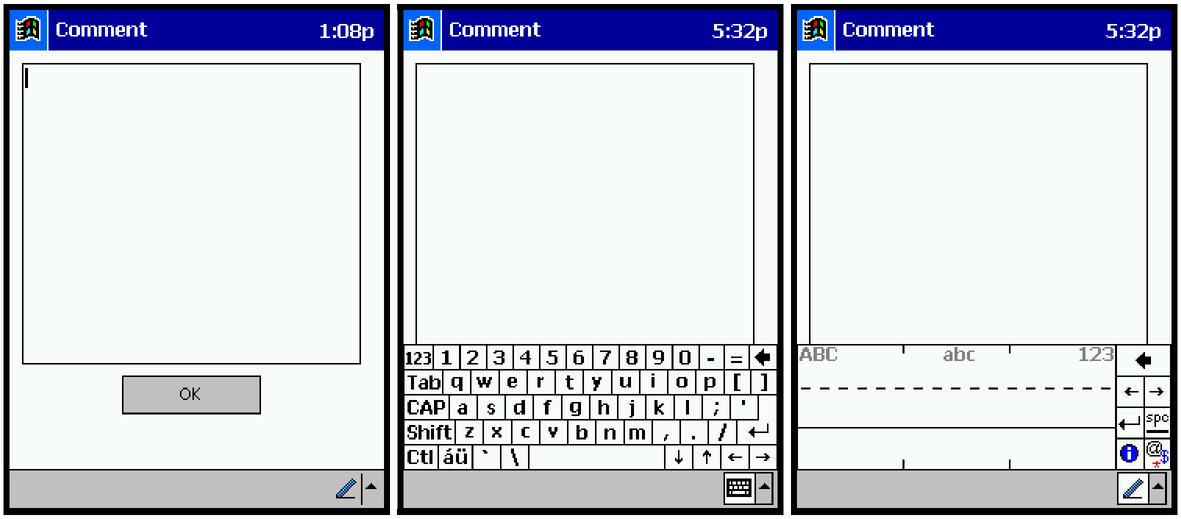


Figure 7. Comment window: with keyboard (lower center) and input panel area (lower left).

is entered, through either method, it is displayed as typed characters in a large text box. Clicking the command button labeled OK will save the comment, time-stamp it, and link it to the active TO/EO/MOP. The instructor is also returned to the Assessment window. The Comment window provides a method by which the instructor can explain and/or augment the selected cause codes measures.

Debriefing Window. The Debriefing window lists all TE's, TO's, EOs, and MOP's in a tree view control. The words complete and incomplete are used to indicate whether the procedure/task described by a MOP was completed or not (see Figure 8).

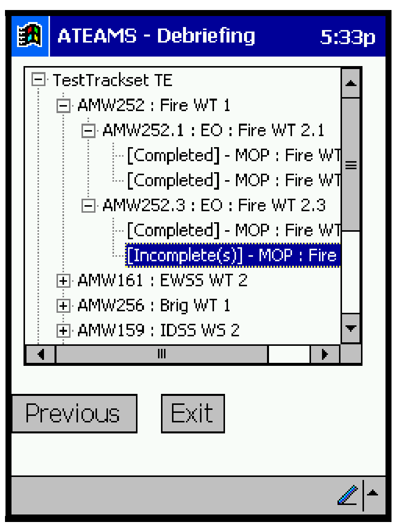


Figure 8. Debriefing window. Top eight lines illustrate a TE, one TO, two EOs, & two MOPs for each EO.

<u>Heuristic Evaluation Procedure</u>

The heuristic design principles applied to the evaluation of the APDA were derived from various sources (Eberts, 1994; Hamel & Clark, 1986; Lynch & Horton, 1999; Mandel, 1997; & Nielsen, 1993). Seven heuristic principles were employed: speak the user's language, minimize users' memory load, provide consistency, prevent errors, provide adequate help and documentation, simplicity, and progressive disclosure (see Appendix A for definitions of these heuristics).

Two human factors experts examined the APDA, evaluating it against the seven guiding principles. Violations of the heuristics were identified and categorized under the three main APDA windows (Prebrief, Collecting Data, and Debriefing) as well as the Collecting Data sub-windows (Timeline, Assessment, Incomplete Cause Codes, and Comment). Note that a heuristic evaluation conducted by one usability expert will reveal approximately 35% of the usability flaws of a given system (Nielsen, 1993). Therefore, between 35% and 70% of APDA usability flaws may have been detected through the current evaluation.

Although it is advisable to fix all identified violations, due to time and financial constraints it is not always feasible. The usability process can be adjusted for this problem by prioritizing each heuristic violation. For example, a priority level of high, medium or low can be assigned to each violation. Violations in the high priority category are issues that are assessed as severely hampering usability; thus, these violations should be resolved through redesigns prior to fielding the system. Medium priority violations should be addressed through redesigns, but are less critical than high priority violations. Low priority violations violate heuristic guidelines, but it is uncertain whether they would impede practical use of the system. Therefore, low priority violations should be addressed, but not if doing so delays the release of the system. The assignment of a priority level to an identified heuristic violation is based on the usability professionals' judgment as to the expected impact each violation will have on user performance. User testing is then employed to confirm or disconfirm the professionals' judgments. User testing sessions can also identify violations missed by the usability professional during the heuristic evaluation process.

User Testing Procedure

The participants in this study had taken part in a previous evaluation of the ATEAMS software (see Ricci, Allen, Reynolds, Daskarolis-Kring, & Hodak, 2001). During that evaluation, the participants were asked to develop a training scenario involving the Combat Systems Training Team and Damage Control Training Team and to perform specified tasks using the APDA software. The objective of the current evaluation was for participants to use and evaluate the GUI, as well as various functionalities, in each APDA window. The evaluation tasks consisted of opening and examining prebriefing material, collecting MOP data, and reviewing the collected MOP data in the Debriefing window (see Appendix B for the instruction set). Additionally, the participants were asked to write a comment on the PDA using the two methods that could be used to input comments. The two methods were the iPAQ's ™ keyboard and Microsoft's® block or letter recognizer software. The participants were asked to verbalize their thoughts about the usability of the PDA and the APDA throughout the evaluation.

APDA Usability Questionnaire. An APDA questionnaire was constructed for this evaluation. This questionnaire was administered immediately after user testing. The questionnaire consisted of four fill-in-the-blank questions, 32 five-point likert-scale questions, and three open-ended questions. Questions 1 – 20 were adapted from the 5NINES Usability Survey by Motorola®. These questions are converted and summed, yielding a score that ranges between 0 – 100 points: 0 equaling very low usability and 100 equaling very high usability. Questions 21 – 32 were adapted from the writings of various usability experts (Eberts, 1994; Hamel & Clark, 1986; Lynch & Horton, 1999; Mandel, 1997; and Nielsen, 1993).

The APDA usability evaluation produced subjective data only. This data consisted of the results of the heuristic evaluation and of the usability survey as well as user's comments. An attempt was made to collect objective data (i.e., number of errors made and elapsed time taken for a given task). However due to equipment limitations, it was quickly determined that this data could not be accurately collected because the researchers did not have access to videotaping equipment. An attempt was made to record time and error data by hand but this proved to be too difficult, given the size of the Pocket PC display.

Results Part 1: Heuristic Violations

A total of 37 heuristic violations were detected. Of these, five were discovered during the user testing sessions with the remainder detected by the Human Factors experts during the heuristic evaluations. A summary of the violations can be found in Table 1. Of the 37 heuristic violations, 25 were categorized as either medium or high priority violations. Note that all of the violations identified under the Prebrief window were rated as low priority violations.

	Users Language	Minimize Memory Load	Consistency	Prevent Errors	Help	Simplicity	Progressive Disclosure	Total: Rows
All Windows		1	2	3	1	2		9
Prebrief Window	1		1	1				3
Data Collection Windows		4	8	4		5	1	22
Debrief Window		1	1			1		3
Total: Columns	1	6	12	8	1	8	1	37

Table 1. Summary of heuristic violations, by window and heuristic.

The following three sections delineate the results of the usability evaluation process. The three sections list the heuristic violations that were common to two or more windows of the APDA application or violations that were specific to a given window (i.e., the Collecting Data windows or the Debriefing window). Within each of these three sections, each violation is number and described. It is then categorized under one of the seven heuristics used during the evaluation and is also assigned a priority level (i.e., High or Medium). User testing data (i.e., user comments and results from the usability questionnaire) that validated the observed heuristic violation, if any, is then given. Finally, redesign recommendations are given. These recommendations are examples of how a violation *may* be addressed, but are not necessarily the *best* solution. That is, the developer may have a more in-depth understanding of the capabilities and limitations of the system being evaluated relative to the usability professional's knowledge of the system. Therefore, the developer may know of a more elegant solution to a given violation (or may know that fixing a violation may not be possible, given system limitations, time constraints and/or budgetary constraints).

<u>Violations Common to Two or More APDA Windows</u>

1. Clicking the Exit button immediately closes the APDA software. No error check message is provided. This increases the possibility that a user may accidentally exit the program.

Heuristic Violation (High): Preventing Errors

User Testing: While looking for a button to back out of the Collecting Data window, one user accidentally exited. The same user accidentally exited from the Debriefing window. His comments include "…oops, where did it go? (accidentally hit Exit). The tendency is that you hit right there first. It would be nice to be able to give confirmation (in wanting to exit)." A second user also accidentally exited from this window. His comments include "That was not good. I hit the wrong exit button I think. Whatever I hit was wrong."

Recommendation: Use the standard Windows® error-checking pop-up window with appropriate text. For example, "Do you want to Exit ATEAMS?" OK/Cancel.

2. The designers of APDA may have purposely employed large command buttons to make it easier for the user to click. The proximity of many buttons could lead to errors produced by mis-clicks (see Figure 9). However, even the smallest buttons used in the APDA software seem to be easy enough to

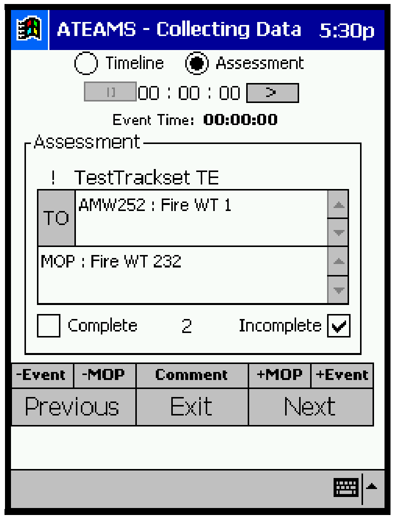


Figure 9. Close proximity of command buttons, like the Exit and Comment buttons, could lead to mis-clicks.

activate; for example, the Timeline/Assessment radio buttons.

Heuristic Violation (Medium): Preventing Errors

User Testing: The mean rating for question 22 of the usability survey, "It was easy to select the checkboxes, buttons, tabs, etc. of ATEAMS HHD", was 4.0. This indicates that the users, on average, agreed with this statement (4.0 = agree with statement). This rating suggests that reducing the size of the command buttons may have no adverse impact on usability since other APDA buttons are already relatively small.

Recommendation: Reduce the size of the Exit/Next/Previous buttons and insert some space between all buttons. This is especially critical for the Exit button to help prevent the user from accidentally clicking this button and exiting the application.

3. APDA provides no access to a help function. This is inconsistent with other applications. A simple help function may be useful to the user, providing guidance when the user needs help with the application.

Heuristic Violation (High): Providing Adequate Help and Documentation and Consistency

User Testing: The mean rating for question 31 of the usability survey, "I need no help when using ATEAMS HHD", was 1.33. This indicates that the users, on average, strongly disagreed with this statement (1.0 = strongly disagree with statement).

Recommendation: Add a help function. Due to limited screen space and memory, the help function will likely need to be simpler than help functions found in standard Windows®-based applications.

4. Users wanted the APDA screens to be as simple and efficient as possible.

Heuristic Violation (High): Simplicity

User Testing: One user, commenting on the method used to scroll through the MOPs/TEs on the Assessment window, stated, "Given that I've developed these MOPs, I'd like to go directly to an MOP. Is it the case that I have to scroll through five or ten of these and the sub-elements beneath? I don't like that. If it's taking a while I'm going to miss a lot of activities that are taking place, missing a lot of behavior that should be observed or recorded." Commenting on the Debriefing window tree view control, another user stated, "Gee, I have to hit plus, plus everywhere".

Recommendation: Adding a drop-down menu may make it easier for the user to find/select the TEs, TOs, EOs, and MOPs from the Assessment window. A drop-down menu may also help the user more quickly locate a given TO, EO, or MOP when using the Debriefing window.

5. On both the Prebrief and Debriefing windows, the user must click on the plus signs used in the tree view control to expand the hierarchy of TEs, TO's, EO's, and MOPs. This is time consuming and may frustrate the user.

Heuristic Violation (Medium): Simplicity

User Testing: One user commented, "Gee, I have to hit plus, plus everywhere."

Recommendation: Include 'Expand all' and 'Collapse all' options. Use a drop-down menu that allows the user to filter the displayed data (e.g., display MOPs only).

6. The timeline in both the Debriefing and Collecting Data windows displays time in scenario time, given in hours, minutes, and seconds (see Figure 10). The time is associated with the point within the scenario that a given TE will begin (this follows the OBT model, i.e., using pre-scripted events in order stimulate and then rate specific areas of trainee performance).

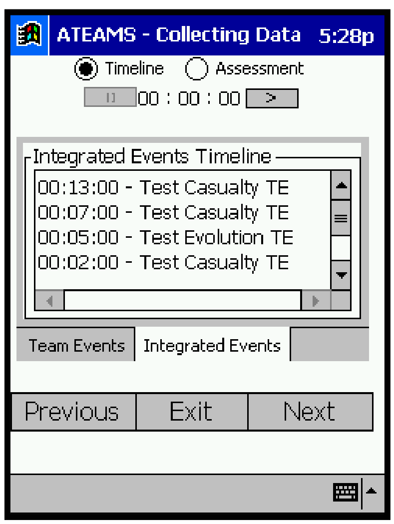


Figure 10. Integrated Events Timeline from Collecting Data window.

Heuristic Violation (Medium): Speak the User's Language

User Testing: One user stated, "It gives me 00:17:00 – that means 17 minutes to analyze and plan? …I'm not used to seeing it like that. 00:17:00 - to me I would read it as seventeen hundred. I would make it plain and simple. Make it 17 minutes."

Recommendation: Use the words 'TE Start Time (hours:minutes:seconds)' as a descriptor of the displayed time.

Violations Relating to the Collecting Data Windows

1. The user should be able to quickly access the data collection functionality, as this is the primary purpose for APDA. When APDA is started, the initial window is the Prebrief window. However, there is no clear indication that the next window is the Collecting Data window, nor is it clear that the user must click the button labeled 'Next' to open that window. This is also the case when a user is finished with data collection and wants to open the Debriefing window. During user testing, it was discovered that several of the users were confused by the textual descriptors or symbols associated with the command buttons (i.e., by the textual descriptors of Next and Previous and by the symbols of + and – associated with the MOP and TE buttons). For example, while on the Assessment window one user clicked the button labeled Previous. This act closed the Assessment window and opened the Prebrief window. The user stated that he thought clicking the Previous button would take him to the previous MOP.

Heuristic Violation (High): Simplicity

User Testing: After exhibiting confusion on how to move forward through the MOPs, a user commented, "Do I hit 'Next' and the next guy comes up…I put complete, and I was saying next, next MOP. Right? That doesn't really make sense." He then tapped the button labeled Next instead of + MOP, which took him to the Debriefing window. Another user also experienced difficulty on how to move from a completed TO to the next TO. His comment was, "Do I go to the next event here?" (He was then directed to the ± MOP/Event buttons).

Recommendation: The Next and Previous buttons can be relabeled to reflect the window that they will open, for example, Prebrief or Collect Data. These labels will need to be abbreviated. To capitalize on a users experience with symbols (e.g., used on a video recorder), an arrow scheme may also be employed for the MOPs and TE's. An example would be as follows: ⇦ MOP. ⇨

2. When the user opens the Collecting Data window, the initial window is the Timeline window. This seemed to confuse the users. During testing, the users often searched for the data collection/assessment portion of the program and had to be provided hints on how to find it.

Heuristic Violation (Medium): Simplicity

Recommendation: Consider making Assessment the default window of the Collecting Data window. If this is not possible, consider making the font for Timeline or Assessment larger or darker (bold).

3. Proper use of screen real estate is critical in any text-intensive PDA application. Two factors affected by PDA screen real estate are readability and comprehension. Readability refers to how easily a user can read displayed text. Comprehension refers to how well a user can comprehend and remember what is/was displayed. Font size, font type, illumination, glare, contrast, and distance from the reader's eye to the text, among other things, affect text readability. In APDA, some of these elements also likely affect reading comprehension. However, comprehension is also affected by the simplicity of the text/sentence and the amount of text that is viewable at once. For example, the text that is displayed in the Debriefing window does not wrap. The result is that the user may have a difficult time in comprehending/remembering long sentences, that is, the user may have to scroll left and right several times to view and comprehend the sentence.

Heuristic Violation (High): Consistency and Minimize User's Memory Load

User Testing: In terms of text scrolling off screen, one user commented, "Well I have to scroll everything, but that's real estate. I'd rather see it wrapped." In addition, the mean rating for question 30 of the usability survey, “I did not have a problem with the length of some of the text, i.e., text that required scrolling”, was 2.33. This indicates that the users, on average, disagreed with this statement (2.0 = disagree with statement).

Recommendation: The best solution may be to wrap all text. Wrapping text provides an additional benefit. That is, by wrapping text the horizontal scroll bar found on the Timeline window can be eliminated. The resulting empty space could be used to enlarge the text box vertically, allowing at least one more line of text to be displayed within the text box.

4. The formatting of the text boxes reduces the amount of space available for text, both on the Timeline and Assessment windows. The formatting between the two windows is also inconsistent (see Figure 11).

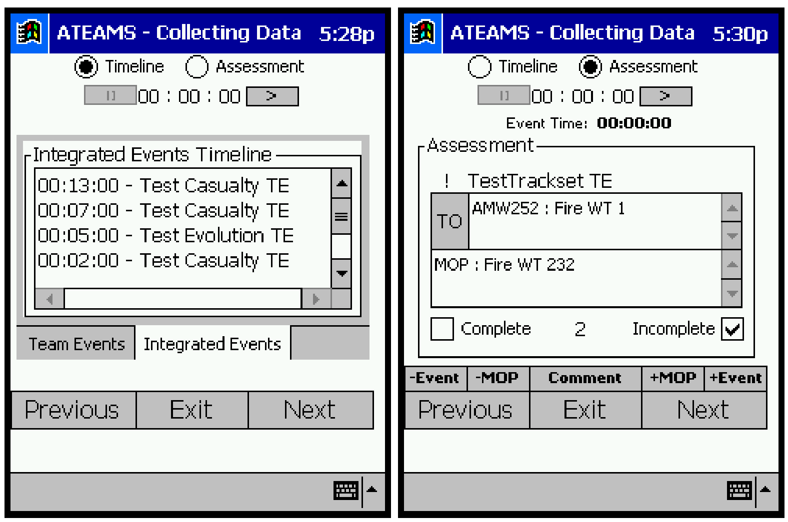


Figure 11. Examples of wasted screen space. Shrink & move tab/command buttons. Reformat textbox area.

Heuristic Violation (High): Consistency and Minimize User’s Memory Load

Recommendation: The gray borders of the text boxes should be removed with the text box enlarged horizontally such that its borders touch the left/right edges of the Pocket PC window. The controls tabs (Team Events, Integrated Events) and command buttons (Previous, Exit, Next) can be reduced in size and moved downward, increasing the amount of vertical space available for the text box. Finally, the textual labels of Team Events, Integrated Events Timeline, and Assessment, which appear just above the text boxes, are redundant (e.g., to open the Assessment window, the user clicks the radio button labeled ‘assessment’). This text can be eliminated, increasing the amount of vertical space available for the text box.

5. The clock control buttons consist of a reset button and a button that toggles between start and pause. No error check message is provided when the reset button is clicked. Thus the user can accidentally click this button and reset the scenario clock to zero.

Heuristic Violation (High): Prevent Errors

Recommendation: Provide an error check message when the reset button is clicked; for example, “Reset scenario clock to zero?” OK/Cancel.

6. The Comment window is composed of an input panel area, located on the lower portion of the Comment window, and a text box display area. The input panel area is used to input text into the Pocket PC; the textbox area displays the text. The user can choose from one of two methods to input text - the user can tap on a virtual keyboard or use Microsoft's® block or letter recognizer software. Once a comment is written, the user saves it by clicking a command button labeled 'OK'. The keyboard seems to be easier to use, but also seems to take more time to input characters. The block/letter recognizer method forces the user to learn how to write letters based on the rules of the software. During the heuristic evaluation, it was noted that this method seemed to lead to more input errors compared to the number of input errors made with the keyboard.

Heuristic Violation (High): Simplicity and Prevent Errors

User Testing: While using the keyboard, a user commented, "It's hard to type." He then began trying to write (erroneously) on the blank comment window. A second user commented about block/letter recognizer method, "Well gee, I wrote on an 'n' and I got a w, l, and v. Let's say that part is not very good, now you're going to train them to write script again and that's not good. When you're putting a training team together you don't have a lot of time". This user was then asked to comment about the keyboard, "The keyboard is OK". He was then asked to comment about the use of digital ink, which had been used for capturing text in other training applications developed at NAVAIR. Digital ink is essentially a drawing application that can capture text exactly as it is written and then store it as a bit map image. "That would be cool, because if I could go to comments and start scribbling straight in there. It would be cool and a lot more useful, instead of me going to this one little keyboard, and having to do the little typing and all that stuff. And it's rapid. You have to realize that as soon as he (a trainee) doesn't do something we might need to stop the drill and I want to make comments about that." A third user commented, "With Palm® you have what they call Graffiti. Some of those characters I have trouble remembering. But I would much prefer being able to write it out".

Recommendation: Rather than forcing the user to learn the format of the block/letter recognizer method or tap on a small keyboard, employ a digital ink method for collecting notes. This method would be the easiest and most accurate method to use. Several software applications have been developed that can be used to create digital notes. For example, Gonna Software® has created a Pocket PC application called PocketStickey, which allows the user to write notes directly on the Pocket PC screen and then link the note to a designated file. Determine if this, or another, application can be used to link a digital ink note to the MOP data collected through APDA. In addition, Microsoft® has created a handwriting software application for Pocket PCs called Microsoft's® Transcriber Version 1.1 for Windows® CE. This application can be tuned based on an individual user's writing style. That is, a user views a representative set of each letter of the alphabet and numbers 0 – 9 and selects, from six different examples of each letter or number, the style that most closely matches his/her own writing style. Investigate this application to determine its error-rate. Military personnel are unlikely to accept error rates higher than 0.5% - 1.0%.

7. To enter a comment, the user selects one of the text entry methods mentioned above. However, all methods cover the OK button of the Comment window, which is used to save the text of a comment (see right two images in Figure 12). To uncover the OK button while in the keyboard mode, the user must click

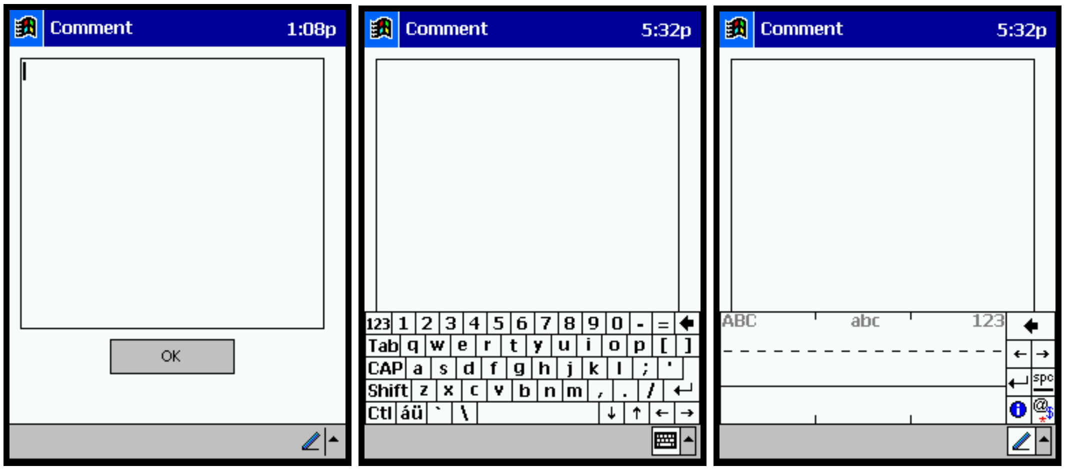


Figure 12. OK command button (left) covered by keyboard (center) and input panel area (right).

on the keyboard icon. There is no indication on screen that this is the method used to minimize the keyboard. If in the block or letter recognizer mode, the user must click a green pen icon to minimize the input panel area. In this case, the background color of the icon changes, depending on the state of the input panel area (gray for inactive, white for active). It is unlikely that the user will recognize this fact. More seriously, if the user accidentally clicks another button – for example, the demo button located to the right of the blue information button - it is quite easy for the user to get lost in the resulting windows and not be able to find his way back to the comment window.

Heuristic Violation (High): Simplicity and Prevent Errors

User Testing: After first using the keyboard, then the block/letter recognizer software, one user had to be shown how to find his way back to the Comment window. That is, he became lost and was unable to save his comment without assistance.

Recommendation: Ensure that the OK button is always displayed. In addition, rename the 'OK' button to 'Save', since that is the operation the user is performing.

8. There is no Cancel button on the Comment window (see Figure 12). This is inconsistent with most Windows®-based applications and could lead to errors during debrief. That is, the user may wish to close the Comment window without leaving a comment. If no comment is entered, the only option the user currently has is to click the button labeled 'OK'. Once the OK button is clicked, a blank comment is inserted into the Debrief database.

Heuristic Violation (High): Consistency and Prevent Errors

Recommendation: Add a Cancel button.

9. The user may wish to edit and/or delete comments before data collection has stopped (see Lyons & Allen, 2000). In APDA, once comments are entered there is no method to recall and edit them. This is inconsistent with other Windows®-based applications.

Heuristic Violation (High): Consistency

Recommendation: Provide a method through which the instructor can recall and edit notes during a training exercise.

Usability Issues Relating to the Debriefing Window

1. The Debriefing window uses a tree view control to display all TE's, TOs, EOS, and MOP's and text to indicate whether the MOP was completed or not. About 40 alphanumeric characters can fit on the window. Unfortunately, the tree view control structure, combined with the length of the MOPs, means that most of the MOP text is off screen and can only be viewed through the use of a horizontal scroll bar. For example, some MOPs are 19 words long. As noted under the Collecting Data section, this formatting may increase the time it takes for the user to view/comprehend the displayed TEs, TOs, EOs, and/or MOPs.

Heuristic Violation (High): Minimize User's Memory Load

User Comment: "Gee, I have to hit plus, plus everywhere." Another user commented, "Well I have to scroll everything, but that's real estate. I'd rather see it wrapped." In addition, the mean rating for question 30 of the usability survey, "I did not have a problem with the length of some of the text, i.e., text that required scrolling", was 2.33. This indicates that the users, on average, disagreed with this statement (2.0 = disagree with statement).

Recommendation: Here we see one of the key limitations of a PDAs, that is, the size of its display screen. Based on the text length that appeared on the APDA, it would seem that scrolling horizontally would tax the users memory load to a greater degree than if the text wrapped. However if the length of the text is long, then wrapping the text may also unduly burden the user's memory load because the tree view control displays all events and objectives, which could number into the hundreds. Ideally no scrolling should be used for an individual MOP. For this to occur in the current configuration, however, the font size would have to be severely reduced, perhaps to an unreadable level. In addition the readability of a given display depends on, among other factors, the amount of stress the user is under, the viewing conditions (light levels, vibration), and quality of the display (Wickens, 1992). No data could be found that directly examined the effects that such factors as text size, scrolling method, stress and/or viewing conditions had on memory or what tradeoffs might be made. Related data would seem to indicate that the use of wrap-around text, drop-down menus, or some other method or combination of methods might be in order (The Windows® Interface Guide, 1995). Additionally, during debriefing the MOP is likely to be the main data point that the instructor would access. If so, it may be useful to separate the TO/EO from the MOP. That is, place the former in a drop-down window, time sequenced, and let the instructor select the TO/EO from this menu (the TO/EO could also be sequenced alphabetically). Then, display the MOPs for the selected TO/EO in a tree view control. It may be useful to provide both options to the instructor (e.g., a 'display all' option), because some instructors may prefer to see the TO/EO/MOP relationships.

2. The MOPs displayed in the Debriefing window have text that indicate whether an MOP was completed or not and, in the latter case, small alpha characters that represent the cause code(s) associated with the reason(s) an MOP was incomplete. However, there is no template to guide the user for deciphering the cause codes. In addition, there is no indication if a comment had been entered for a given MOP nor is there a method for the user to access the comment associated with an MOP (see Figure 13).

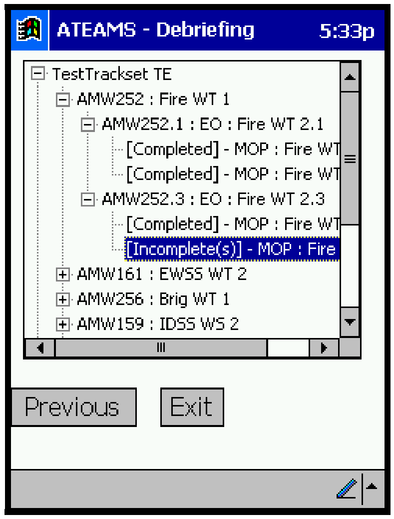


Figure 13. Debriefing window with complete/incomplete text and cause code (s in highlighted example).

Heuristic Violation (High): Consistency and Simplicity

User Comment: One user commented, “I guess would want to have some indication as to where I've made a comment - without having to look through every one of them.” A second user commented, “Say there is one incomplete, where do I see the comment?” (explained that he can't from here). "Why not? I went through all that process of writing stuff in so that in debrief I can see it. No, it should have a link right here. Put a little tab, something so that the comment opens up. If you say incomplete and there are twelve people, they want to know why it's incomplete. How are you going to learn if you don't know why it wasn't complete? To me that's really big fault there." A third user commented, "Let me see if I can find one that has an incomplete. That's all it's going to tell me? Where are my personal notes? That's got to be in there. Otherwise, I just might not remember, especially if there are a number of incomplete."

Recommendation: Provide a cause code template to the user (perhaps through a help button). Add a comment code to the MOP so the user will know which MOP has a comment associated with it. Provide a method through which the user can display his comments; for example, by tapping on the MOP.

Results Part 2: APDA Usability Survey

An APDA usability survey was constructed for the evaluation. It was administered immediately after user testing. The results of the survey can be found in Table 2.

Question (based on a five-point likert scale: 1 = Completely Disagree; 3 = Neutral; 5 = Completely Agree).	**Mean**
Years of Experience in Training Community	20.00
1. ATEAMS HHD is difficult to use.	2.00
2. I have a good understanding of how ATEAMS HHD features are organized.	3.33
3. I would not want to use ATEAMS HHD every day.	2.33
4. The layout of the buttons made sense.	3.00
5. The physical layout of ATEAMS HHD make it difficult to use.	3.33
6. The steps to complete tasks followed a logical sequence.	2.67
7. I had to tap too hard on the buttons to operate them.	1.33
8. The icons and the graphics on the buttons/keys were clear.	3.00
9. I was confused by the organization of information on the display.	3.33
10. It was easy to learn how to use ATEAMS HHD.	3.33

11. I could not relate the words on the screen to the tasks.	3.00
12. It was easy to explore ATEAMS HHD features by trial and error.	3.67
Question (based on a five-point likert scale: 1 = Completely Disagree; 3 = Neutral; 5 = Completely Agree).	**Mean**
13. It took too many steps to complete tasks.	3.00
14. It was easy to read the words on the screen.	4.33
15. I could not tell when I had completed an action correctly.	3.67
16. It was easy to correct mistakes.	2.67
17. I was confused by the terms on the screen.	3.33
18. It was easy to find the features I wanted.	2.67
19. I often felt lost and did not know how to proceed.	3.00
20. The display size was not too small.	4.00
Motorola 5NINES Score: Questions 1-20 (0 = very low usability; 100 = very high usability)	55.67
21. The information provided on the prebrief screen would be useful during a training exercise.	3.67
22. It was easy to select the checkboxes, buttons, tabs, etc. of ATEAMS HHD.	4.00
23. The terms used on ATEAMS HHD were understandable, i.e., ATEAMS HHD 'spoke my language'.	
24. I found the interface of ATEAMS HHD to be highly intuitive.	3.00
25. It was easy to navigate through ATEAMS HHD.	3.00
26. It was not easy to recover from mistakes made with ATEAMS HHD.	3.67
27. The information provided on the debrief screen was useful.	3.67
28. The performance rating system of the data collection screen made sense.	2.33
29. Useful feedback was provided by the computer whenever I completed an ATEAMS HHD task.	2.67
30. I did not have a problem with the length of some of the text, i.e., text that required scrolling.	2.33
31. I needed no help when using ATEAMS HHD.	1.33
32. ATEAMS HHD does not need much improvement.	2.00

Table 2. APDA Usability Survey Results.

An examination of the data found in Table 2 seems to indicate that the users felt that the APDA software had poor usability. This is indicated by the 5NINES summary score of 55.67 (out of a possible 100) as well as by the answer to question 32, "ATEAMS HHD does not need much improvement", which garnished a mean rating of 2.00 indicating that the user disagreed with this statement. However, they also disagreed with the statement that the ATEAMS HHD was difficult to use. It may be that the users found the *iPAQ™ hardware* easy to use *physically* (e.g., button clicks, simplicity of text) but the *process* of using the *APDA software* difficult (e.g., moving from one screen to another, entering a comment). For example, questions 1, 6, and 16 are process questions and indicate that the users experienced some difficulty with the APDA process associated with each question. Alternatively, questions 4, 7, and 8 are questions relating to the physical features of iPAQ™ H3650 and seem to indicate that the users did not have much difficulty with the physical feature associated with each question. The difficulty with the APDA process was born out by user comments:

1. "When you are grading an exercise you are not going to have a lot of time really to switching back and forth."
2. "Everything needs to be as expeditious as possible."
3. "Given that I've developed these MOPs, I'd like to go directly to an MOP that - is it the case that I have to scroll through five or ten of these and the sub-elements beneath."
4. "I'm sort of lost on how to get to the debrief section."
5. "I guess would want to have some indication as to where I've made a comment - without having to look through every one of them."
6. "Do I hit 'Next' and the next guy comes up? Right? That doesn't really make sense."

In addition to the process questions, the users also did not feel that the APDA performance rating system made sense (question 28). This was confirmed by one user's comment, "These gross classifications, I'm not sure if they would be much value to me." Whether more complex performance assessment methodologies can be placed on a PDA with its limited screen real estate remains to be seen.

Summary

Overall, 37 violations were identified during the heuristic evaluation and user testing sessions. The breakdown of these violations can be found in Table 3.

Violated Heuristic	Number of Occurrences
Speak the user's language	1
Minimize users' memory load	6
Consistency	12
Prevent errors	8
Provide adequate help	1
Simplicity	8
Progressive disclosure	1

Table 3. Breakdown of Heuristic Violations.

As can be seen in Table 3, the consistency heuristic was most frequently violated. This is not surprising, given that APDA is prototype software. This is not to say that these violations are not important and should not be fixed. For example, one violation that fell under the consistency heuristic was that comments were not accessible from the Debriefing screen. This was a major source of frustration for the users and would severely reduce the utility of APDA. The number of violations that fell under the simplicity heuristic indicates that the APDA process should be simplified. The number of violations that occurred under the memory load and prevent errors heuristics also reinforces this notion.

Several correctable errors occurred during the evaluation, two of which should be corrected immediately. First, the user can accidentally exit the system without receiving an error-checking message. Second, the user can accidentally reset the scenario clock without an error-checking message appearing. The users also expressed concern about the errors committed when using two of the methods for inputting comments via the iPAQ™ (i.e., the scripting and electronic keyboard methods). The use of digital ink would correct this problem. Several software applications have been developed that can be used to create digital notes. For example, Gonna Software® has created a Pocket PC application called PocketStickey, which allows the user to write notes directly on the Pocket PC screen and then link the note to a designated file. Seiko Instruments USA Inc. has developed a unique input method that can be used for hand-helds, laptops, and PCs. It is called the InkLink™ handwriting system. To use this system, the user attaches a large clip, dubbed a data clip, to a pad of ordinary paper. The InkLink™ system tracks the movement of an electronic pen and sends this data to the computer device, which faithfully displays what the user has written and/or drawn.

Presentation and readability of displayed text were also areas of concern, both for the users and evaluators. Although the users indicated that they thought that the size of the text was adequate, they also noted that they had a problem with the length of the text (i.e., text that required scrolling). Because text length is affected by both the size of a font and its type, a font size and type/number of characters tradeoff occurs for PDAs. That is, the size and type of font used affects the number of characters that will fit in a given text box. Currently between 30 and 34 characters (including blank spaces) can fit into the Team Events Timeline and Assessment text windows before the text either disappears off screen (true for the Team Events Timeline window) or wraps (true for the Assessment window). The most common font sizes used in newspaper text ranges between nine and 12 points (Sanders & McCormick, 1987). This size can be used as a baseline when trying to apply standards to a computer display. For such displays, the font should subtend a visual angle of at least 12 minutes of arc (Sanders & McCormick, 1987). This would mean that at a viewing distance of 18 inches, the font size should be at least six-points. Such a small font size is in agreement with the user's assessment of the readability of the text on the iPAQ™, as indicated by the usability survey. However other factors, such as lighting conditions, vibration, criticality of information, visual acuity of the user, and amount of stress the user is under can also affect the readability of a display (Wickens, 1992). Under these circumstances, the font size should probably be increased beyond the minimum six-point font size. Some investigators indicate that legibility and readability would be increased with larger font sizes, sizes that subtend up to 25 minutes of visual angle (Wickens, 1992). At a viewing distance of 18", a 25-minute viewing angle translates into a 13-point font size. Based on the above data, it seems reasonable to suggest that the APDA font size match the font size range found on newspaper print (i.e., from a minimum of nine to a maximum of 12-points).

The key to the above is to ensure that all information displayed on APDA, especially the most important information, is readable and easily comprehensible. For the Timeline and Assessment windows, the most important information is found within each text box (e.g., the timeline of TEs and the MOPs, respectively). Therefore, the size of the text boxes should be maximized. The size of this display area also affects how much scrolling a user must perform. Unfortunately, in text-heavy applications, a PDA may lead to the type of frustration expressed by the participants in this evaluation. Therefore it is recommended that the Navy test the usability of each class of computer device (e.g., PDA and pen tablet computer) onboard ship during live training exercises with instructors, before committing to a given type or brand of device. Testing under real training conditions is critical in any evaluation. Such tests could reveal flaws in the design of the application and/or equipment that would not be found under more sterile testing conditions.

Other recommendations regarding usability issues have been outlined throughout the document. Specific recommendations include:

1. Error-checking when a user clicks the exit button or reset clock button.
2. Allowing the user easy access to comments.
3. Simplifying the APDA process. One way to do this would be to use clear textual descriptors to help indicate how a user moves from one screen to another.
4. Reducing the need for scrolling (e.g., by providing wrap-around text).
5. Reducing the size of the buttons. This should help reduce the potential for accidental activation and also provide additional space for the text box display area.
6. Using digital ink for comments.

Several of the users indicated that the information collected via the data collection screen may not be of much value. The simple data collected through APDA may be a limitation of the size of Pocket PC screens. It is recommended that a task analysis be conducted to define the type of data that shipboard instructors need. If an application with a more complex interface, or that is more text-intensive in nature, is required then the U.S. Navy might consider investing in pen tablet computers for its instructor data-collection requirements.

Implementing the APDA Redesign Recommendations

The developers of APDA received a comprehensive version of the data and results presented in this paper. The first step taken by the developers was to consolidate the identified design issues, isolating those situations in which solving one design issue would eliminate other design issues. The design violations were organized by APDA window or affected windows and then by the related heuristic. Doing so provided a framework that was used to reference specific design issues and to identify actions to be taken to correct the issue (see Appendix C).

Once the recommended modifications were categorized and actions needed to implement the modifications identified, a prioritization process was initiated. Three categories were used to help prioritize a given redesign recommendation.

1) How the modification would affect the user's understanding of how to use the APDA
2) The required level of effort to implement a modification.
3) How the modification would affect the functioning of the APDA

Two SMEs assessed the recommendations in the framework, determining which modification(s) would lead to the most significant improvement(s) for the APDA end-user. The SMEs rated each modification as important, nice to have, or not important. In parallel to the SME assessments, software designers evaluated the level of effort that a given action (i.e., implementing the modification) would require. Embedded within the level of effort assessment was an evaluation of how an action would affect the functioning of the APDA application. The factors that were considered for assigning level of effort included time, cost, and supportability. Based on the relative level of effort required, the actions were then classified as easy, medium, or hard. An easy action would require minimal time and/or research effort to accomplish. A hard action might require writing new controls for an APDA function, conducting an extensive search for existing software solutions, or identifying an issue that could only be resolved by significantly reorganizing how the ATEAMS database was arranged. A medium action would fall somewhere between an easy and a hard level of effort. The results of the SME and developer assessment efforts was a prioritization matrix, with each action falling somewhere on the important – not important and easy – difficult continuums (see Appendix D).

The final step of this process was to develop an implementation schedule. Each of the actions was reviewed based on the prioritization matrix and available resources to determine which action would be implemented first. From this review, each action was prioritized as a Priority 1, 2, 3, or 4 action. All Priority 1 actions were to be completed prior to beginning Priority 2 actions, all Priority 2 before Priority 3, and so on. This prioritization scheme enabled Sonalysts to efficiently allocate programmer time to the APDA redesign (see Appendix D).

It should be noted that several of the identified redesign recommendations were treated as special cases. For example, the recommendation to provide help functionality and documentation was an effort that was well beyond the funding and time requirements available for implementation and thus was treated as a separate project. Several other actions were also classified as "not to be implemented in the prototype". These actions were so classified due to certain functionalities that were not supported by either the APDA operating system or the programming environment or due to resource requirements that would not be available during the project performance period (e.g., evaluation by actual end-users. See Appendix D).

Implementation Examples

The prioritization scheme meant that modifications categorized as easy-to-implement and important would be given a priority 1 rating. This scheme is reasonable, but could mean that critical errors would not be corrected. For example, the usability evaluators determined that accidentally exiting APDA was a high priority design violation. Had the developers and SMEs rated this as a medium-level effort and a nice-to-have or as a high-level effort, but important, it would not have been given a priority 1 rating. Yet, this violation was critical, given that the user would have to restart the application, losing valuable assessment time and potentially missing key observations.

What follows are examples of APDA windows, both before and after recommendation(s) from the usability evaluation were implemented. This section applies an abbreviated format to the one used above. First the violation that was addressed is numbered and described. The violations are grouped under violations that were common to two or more windows, to the Collecting Data windows or to the Debriefing window. The redesign recommendation(s) to the violation(s) is (are) then noted. The original window, and the redesigned window, is then shown.

Violations Common to Two or More APDA Windows and Implemented Solutions

1. Clicking the Exit button immediately closes the APDA software. No error check message is provided. This increases the possibility that a user may accidentally exit the program.

Solution: An error-checking window now appears when the Exit button is clicked (see Figure 14).

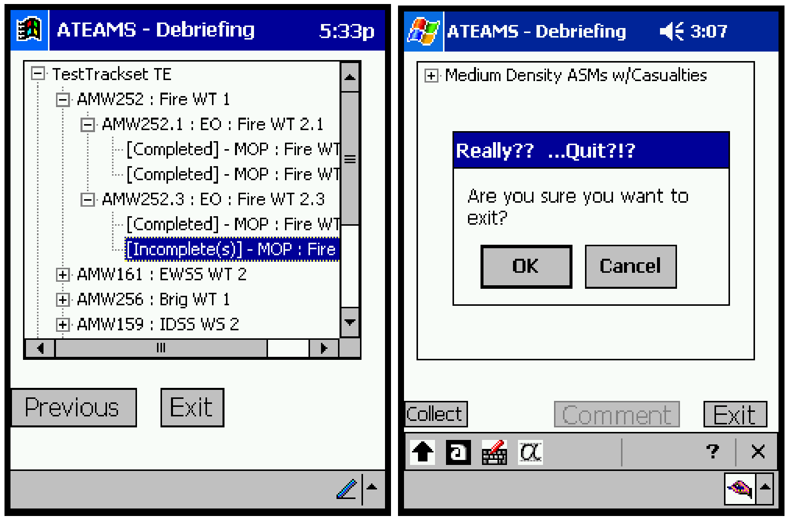


Figure 14. Original (left) and redesign. Note error-checking message.

2. To make it easier for the user to click, the designers of APDA may have purposely employed large command buttons. The size and location of many of the buttons could lead to errors produced by mis-clicks.

Solution: Command buttons were resized, moved, and/or separated in space (see Figure 15).

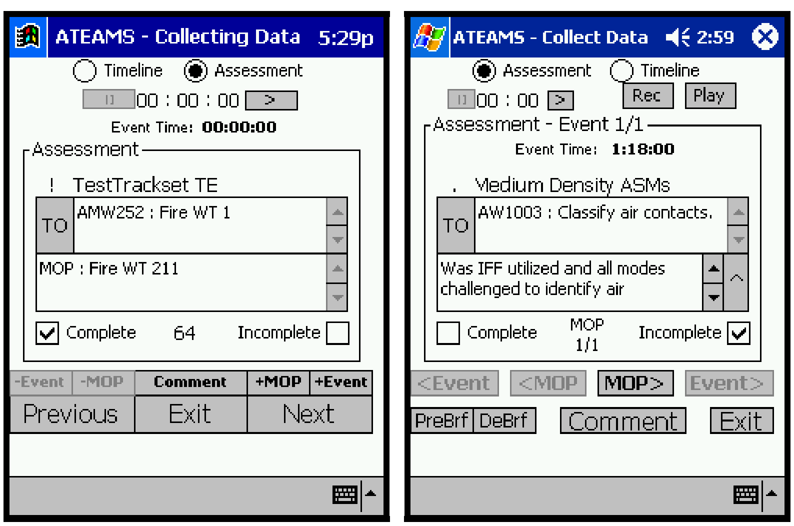


Figure 15. Command button location increases potential for mis-clicks (left). Note size/spacing changes on right.

Violations from the Collecting Data Windows and Implemented Solutions

1. When APDA is started, the initial window is Prebrief. However, there is no clear indication that the next window is Collecting Data, nor is it clear that the user must click the 'Next' button to open that window. This

is also the case when a user is finished collecting data and wants to open the Debriefing window. During user testing, it was discovered that several of the users were confused by the textual or symbolic descriptors associated with the command buttons (i.e., by the textual descriptors of Next and Previous and by the symbolic descriptors of + and – associated with the MOP and TE buttons). For example, while on the Assessment window one user clicked the button labeled Previous. This act closed the Assessment window and opened the Prebrief window. The user stated that he thought clicking the Previous button would take him to the previous MOP.

Solution: The navigation command buttons were renamed, reflecting the name of the window that they open when clicked. For example, on the Collecting Data window, the Previous and Next buttons were relabeled to PreBrf and DeBrf, respectively. Additionally, the + and – symbols found on the MOP and Event buttons were replaced by VCR-like control symbols. The idea is that these symbols will tap into a users experience with symbols found on VCRs, and other similar devices, making it easier for them to interpret the meaning associated with these symbols (move forward, move back). These changes can be seen in the right window in Figure 15. Also, when the user moves from the Prebrief window to the Collecting Data window, the Assessment window is now the first window to open. Previously, the Timeline window was the default window.

2. To enter a comment, the user selects one of the text entry methods. However, all methods cover the OK button of the Comment window. The user also cannot exit out of this window without saving a blank comment.

Solution: The command button was moved above the display area. Additionally, the command button was renamed Save and a Cancel button was added (see Figure 16).

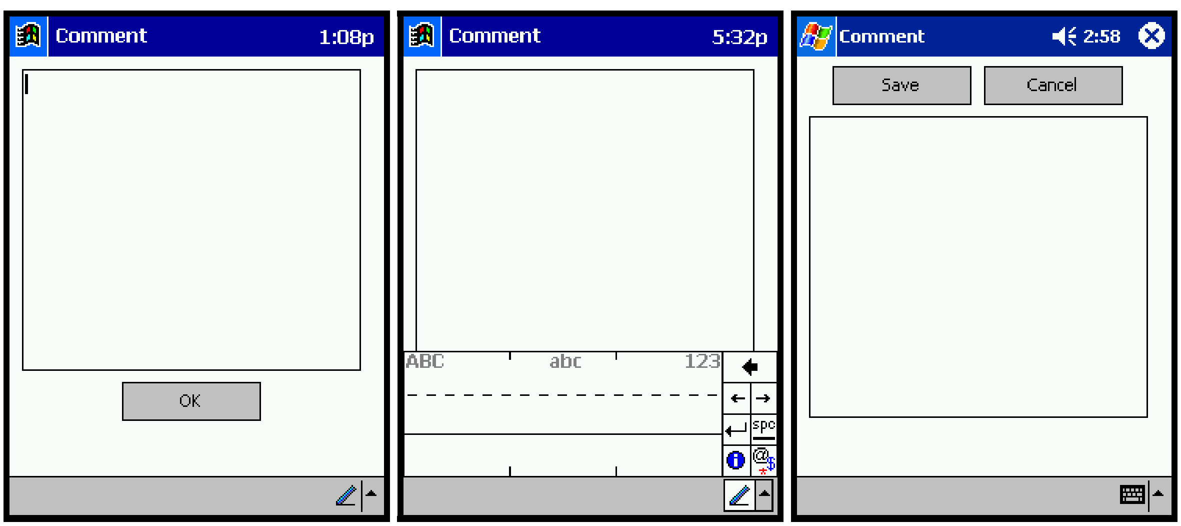


Figure 16. Command button (left), hidden (center). Solution: moved above display area (right).

3. The user may wish to edit and/or delete comments before data collection has stopped, but APDA does not provide this capability.

Solution: A comment command button was added to all windows. Additionally, the user can edit comments during data collection (see Figure 17).

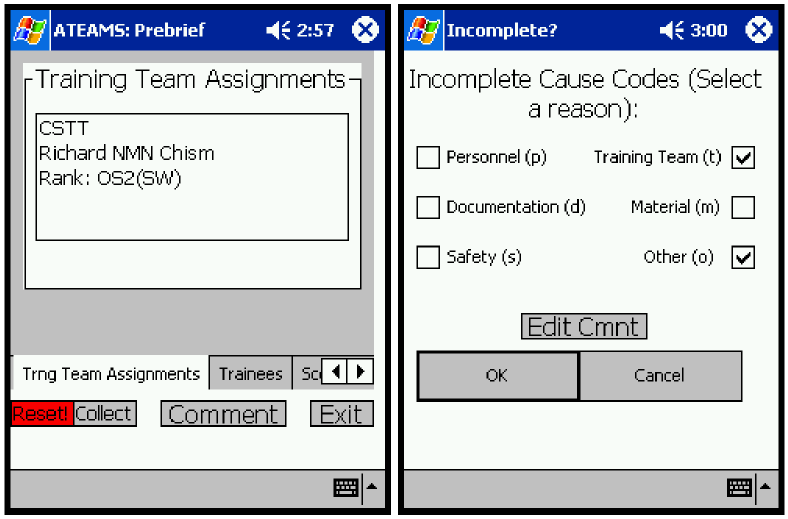


Figure 17. Comment link added to Prebrief (left). Comment can be edited in Incomplete? window.

<u>Violations from the Debriefing Window and Implemented Solutions</u>

1. The text that is displayed in the Debriefing window does not wrap. The result is that the user may have a difficult time in either comprehending/remembering long sentences or may take more time to read the text compared to the time it would take to read text that was viewable at once.

Solution: The user can double-tap on a given MOP in the Debriefing window, which will then open a separate window that displays the MOP in a large font (see Figure 18).

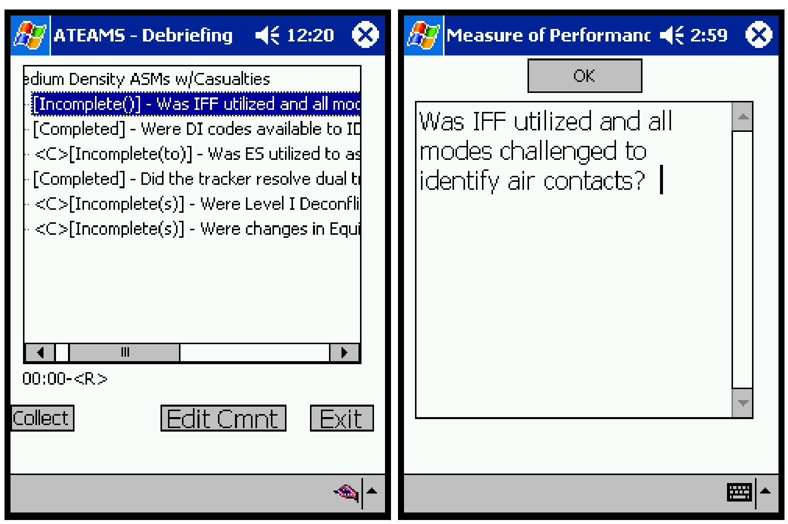


Figure 18. Double-tapping on an MOP listed in the Debriefing window now opens a display window.

3. The Debriefing window displays whether an MOP was complete or incomplete, with cause codes listed with the latter. However, there is no indication if a comment is attached to an MOP.

Solution: A comment code has been inserted into the MOP list. The code is displayed as <C> (see Figure 19).

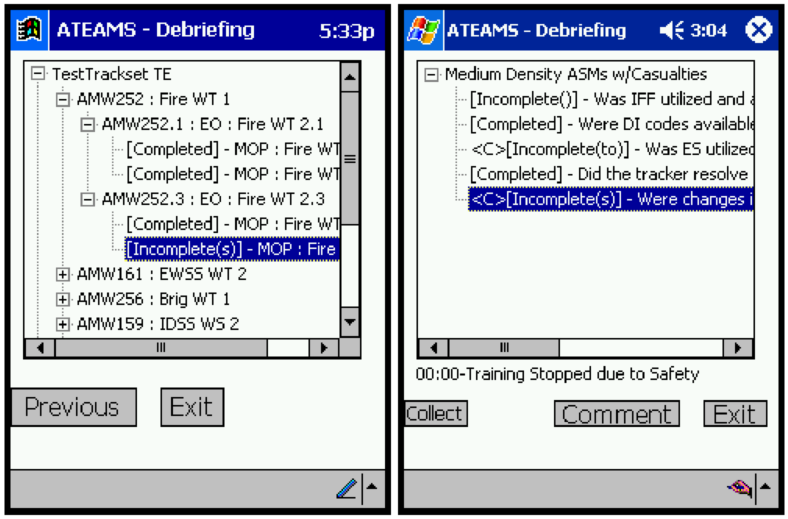


Figure 19. Initial version of Debriefing window (left). Redesign with comment code <C> (right).

Current and Future Efforts

The APDA usability evaluation was a preliminary evaluation of prototype software. The results have been used to drive the redesign of the application's windows. One key to increasing the usability of a product is iteration, meaning the system should go through continual refinements until pre-established acceptability criteria have been met. Therefore, the redesigned APDA will also be subjected to a usability evaluation. The usability process is also meant to be inclusive; meaning the usability professional, the system developers and the end-users should be involved from design conception to completion. So far, the APDA process has been iterative. However, the process also needs to be inclusive. Therefore, an attempt will be made to contact end-users and involve them in the design process.

There are several new efforts underway towards implementing data collection capabilities through PDAs. Two examples follow. Each of these efforts has applied knowledge gained in the APDA usability evaluation process to the design of the newer applications.

<u>Coalition Readiness and Exercise Management System (CReaMS)</u>

The Coalition Readiness and Exercise Management System (CreaMS) is a project designed to develop efficient and effective training situations for individuals and teams that are distributed around the globe. The targeted training audience is a coalition Naval Task Group and its members. Participants include Australia, the Netherlands, and the United States. Observations of past exercises identified areas of coalition operations that could be significantly improved with the employment of proper training. The CreaMS effort identified the need for an effective debrief between coalition partners at the conclusion of an exercise (i.e., timely and meaningful interactions between coalition partners that would review the strengths and weaknesses of each partner's efforts in the exercise).

One of the requirements for an effective exercise debrief in this context is that debriefing data be available to all coalition partners. ATEAMS and the APDA were used to provide exercise coordinators a means by which they could rapidly develop a training scenario and create a data collection plan tailored to the particular watchstations that an instructor would be observing. They also were used to quickly distribute the data collection plan and to compile the collected data fast enough for a timely debrief. The developers of the CreaMS APDA applied lessons learned from the APDA usability evaluation to the CreaMS application, likely reducing development time that might have otherwise been needed for redesign efforts.

Sonalysts has also created an ATEAMS-independent hand-held data collection tool that is currently being used with revised CReaMS databases. Antidotal data suggests that incorporating the results of the APDA usability process early in the design phase produced a product that was easy for the end-user to learn.

Battle Stations 21

The Battle Stations 21 project is designed to produce a capstone, 24-hour Navy Basic Training experience. All recruits must complete Battle Stations 21 prior to graduating from Boot Camp. Battle Stations 21 will present realistic scenarios that challenge the recruit's determination, endurance, and core Navy values of honor, courage, and commitment. During Battle Stations 21 recruits will be provided an opportunity to demonstrate a variety of basic Navy skills acquired during the previous seven weeks in recruit training. These skills cut across a variety of domains from protocol and etiquette to Seamanship and Damage Control. An overarching goal of the events and scenarios of Battle Station 21 is to foster a team environment in which recruits must work together to complete exercises or overcome obstacles.

Battle Stations 21 will be an intensive experience that introduces a number of challenges for the instructor. Instructors must provide a realistic training experience in a training environment that is highly constrained, in terms of both time and space, and that also has a large student throughput requirement. To achieve the goal of Battle Stations 21, a state-of–the-art Training Management System will be developed. A hand-held data collection tool similar to the APDA will be part of the Training Management System. The following sections describe how the hand-held device will be implemented within the Training Management System.

Data Collection Module. The Training Management System will be used to create the exercises' data collection packages, which will then be downloaded to a hand-held device. The hand-held device's software will enable instructors to collect performance data and to make observations that are related to safety, the scenario story, and scenario integrity.

Debrief Support Module. Once performance data is collected, the hand-held device will be used to present debriefing material (e.g., the hand-held device will be connected to a projector that will project the debriefing material onto a screen). This requirement means that the software must allow instructors to quickly summarize and prepare data for debrief. In a similar vein, debriefing support software will also be implemented on the hand-held device. Instructors can use the debriefing support software to guide them through a pedagogically sound debrief.

Instructor Communication Module. With the advent of reliable and affordable wireless networks, the hand-held device's software will also support exchange of information between instructors and between instructors and exercise operators/controllers. This software must allow personnel to communicate with minimum effort. Ease of use is a key concern of instructors, whose primary focus must be on what is happening with the recruits during a training exercise.

Conclusion

To test the feasibility of using a PDA for data collection during shipboard training exercises, researchers at Sonalysts, Inc. developed a prototype data collection application ATEAMS PDA (APDA). The initial application could be used to synch the Afloat Training Exercise and Management System (ATEAMS), hosted on a PC, with a PDA. Once synched, training relevant data could be downloaded from ATEAMS to the PDA. Researchers at NAVAIR Orlando Training Systems Division subjected the APDA to a usability evaluation. Usability evaluations can reveal critical and non-critical design flaws in hardware and software and, therefore, was chosen as one method for evaluating the APDA application. This evaluation consisted of heuristic evaluations, user testing sessions, and redesign recommendations. Three users that had taken part in the ATEAMS software evaluation were asked to perform specific tasks using the APDA software. The user testing sessions were designed to validate the heuristic violations or find problem areas not identified by the usability evaluators.

Thirty-seven violations were identified during the heuristic evaluation and user testing sessions. Redesign recommendations were provided to the developers of APDA. The developer used the APDA report to prioritize the heuristic violations. Two SMEs determined which modification(s) would lead to the most significant improvement for the APDA end-user. The SMEs rated each modification as important, nice to have, or not important. In parallel to the SME assessments, software designers evaluated the level of effort that a given action (i.e., implementing the modification) would require. The factors that were considered for assigning level of effort included time, cost, and supportability. Based on the relative level of effort required, the actions were then classified as easy, medium, or hard. The result of the evaluation process was a prioritization matrix. The data within this matrix drove the redesign effort.

The APDA evaluation demonstrated that the usability process could be applied to small form factors like Pocket PCs. The evaluation also indicated the need to develop tools that could be used to collect objective data such as time and error rates, when evaluating such small form factors. Videotaping equipment could be employed only if the PDA were mounted, which would introduce artificiality in the evaluation because users generally hold PDAs. However, Noldus Information Technology® has developed a wireless camera that can be mounted on a PDA. This device is relatively unobtrusive and would be well suited for videotaping such devices (Noldus Information Technology, 2003). There is a need for a PDA software application that can collect time and button-click data, which could then be analyzed for error events. Such applications currently exist for standard PCs. The Group for Interface Research at the University of California at Berkley has developed an application that records button clicks for web-enabled hand-held devices called WebQuilt. This application records the web links that a user clicks on, the web pages visited, as well as time spent on a page. However, it cannot record other onscreen clicks (e.g., when a user clicks on a scroll bar) or user comments (Waterson, Matthews, & Landay, 2002).

The APDA evaluation revealed limitations in PDAs when they are applied to U. S. Naval data collection efforts. The primary limitation is the size of the display screen. Because of this limitation, a PDA may not provide the best solution for text-intensive data collection applications. A secondary limitation relates to the input methods described in the paper. The Pocket PC does not provide an acceptable solution when quick and accurate note taking is required. Microsoft® has introduced handwriting recognition software, both for Pocket PCs and pen tablet computers. Preliminary evaluations of both applications are not encouraging. The former does not seem to provide the accuracy needed for training environments. The latter's method of converting handwritten notes to text seems to be extremely labor intensive and non-intuitive. Although the recommended method may tax the storage capacity of a Pocket PC, digital ink may be the best solution for electronic note taking for a PDA in a shipboard training environment.

Appendix A: Heuristic Definitions

1. Speak the user's language (rather than computerese). Text displayed to the user should be expressed in words, phrases, and concepts familiar to the user. Input and output must conform to population stereotypes.
2. Minimizing users' memory load. The user should not have to remember information from one part of an application to another. Instructions for use of the system should be visible or easily retrievable. Methods of chunking and focusing information aid in reducing mental workload.
3. Consistency. Users should not have to wonder whether different words, situations, or actions mean the same thing. If the user develops a correct mental model of the system, there will be a dramatic reduction in cognitive processing expended on understanding how the system works. One of the major benefits of consistency is that users can transfer their knowledge and learning to a new program if it is consistent with other programs they already use.
4. Preventing errors. Even better than good error messages is a careful design that prevents a problem from occurring in the first place.
5. Providing adequate help and documentation. Even though it is better if the system can be used without documentation, it may be necessary (and is advisable to accommodate all users types) to provide help and documentation. Any such information should be easy to search, be focused on the user's task, list concrete steps to be carried out, and not be too large.
6. Simplicity User's are not impressed with complexity that seems gratuitous, especially those who may be depending on the application for timely and accurate work-related information.
7. Progressive Disclosure. Users should not be overwhelmed by what they can do in a product. You don't need to show users all of the functions the product offers. The best way to teach and guide users is to show them what they need, when they need it, and where they want it. This is the concept of progressive disclosure. New technology such as wizards and assistants use progressive disclosure to guide users through common tasks. Wizards guide users through steps in a progressive manner where each step is simple and meaningful for even casual users.

Appendix B: APDA Scenario Instructions

Last time we worked together, we asked you to develop a training scenario using the ATEAMS software. The training teams in this scenario were the Combat System Training Team (CSTT) and the Damage Control Training Team (DCTT). The Mission Area was Under Sea Warfare (USW). The idea was that you were onboard ship, developing a training scenario to be used for shipboard training. After completing the scenario, the next step would be to download the trainee performance evaluation sheets provided by ATEAMS to a Personal Digital Assistant (PDA). We have done this for you. This is the purpose of the current evaluation, i.e., you are now going to use and evaluate the data collection portion of ATEAMS that has been loaded onto the PDA.

As you use the PDA and the ATEAMS software, we again ask that you "think aloud" so that we can accurately interpret misunderstandings, opinions, and expectations you may have related to ATEAMS. We may prompt you at times for such information. During this exercise we will be unable to assist you unless absolutely necessary. You can refer to the information below while using the PDA. If you have any questions, please ask them now. Do not start ATEAMS until the experimenter gives you the go-ahead.

Tasks:

1) Turn on the PDA and start ATEAMS HHD (ATEAMS Hand-Held Device).
2) Using the provided hand-held electronic device containing ATEAMS HHD:
    a. Confirm the Training Team Assignment as CSTT
    b. Confirm the Trainee(s) as John Hodak.
3) Look at the Timelines and Scenario Summary and share your thoughts on what information is provided and how it is provided.
    a. The Lessons Learned, Safety, and ROE screens are empty, but feel free to explore and comment on those screens.
4) Review the Objectives and share your thoughts on what information is provided and how it is provided.
5) Document the given Measures of Performance (MOPs) for the named trainee. You will be acting as an instructor assessing the trainee, John Hodak, on the listed MOPs. He is a trainee for the watchstation Time/Brg/Time Freq Plot. When you begin the assessment,
    a. Start the timing clock.
    b. Rate the listed MOPs as complete or incomplete (see attached form).
        i. If the MOP is incomplete, document the Cause Code. All the information you need to enter is listed in the table below.
6) While working with ATEAMS HHD, share your thoughts on what information is provided and how it is provided.

Measures of Performance (MOPs) USW		Complete	Incomplete: Cause Code(s)
201	*Analyze and Plan for a USW Mission or Task*		
	Incidents of conflict not resolved.	C	
	Time to prepare/promulgate plan.		T (training team)
205	*Search, Detect, Localize, and Track*		
	Range from own unit/forces contacts(s) detected versus predicted.		P (personnel, O (other)
	Time to fix contact's position.	C	
101	*Detect, Classify, and Track Subsurface Contacts.*		
	Was SITSUM display monitored?	C	
	Were assigned verniers replicated?		S (safety)
103	*Control Aircraft in a USWW Role.*		
	Was classification of the target updated?		M (material)
206	*Classify and ID Subsurface Contacts.*		
	Range from own unit/force contact classified/identified.		D (documentation) S (safety)
	Time to classify/identify contacts.	C	
207	*Engage to Achieve Mission*		
	Time to develop firing solution(s) or engagement plan.	C	
	Time to conduct attack.		T (training team M (material) O (other)
102	*Engage Subsurface Threat with Antisubmarine*		
	Were re-attack procedures conducted?		D (documentation)

6) When finished, submit the data for debriefing and look over the debriefing material provided for the assessed objectives. Again, please share your thoughts on what information is provided and how it is provided.

Appendix C: Prioritized Actions to be Taken on Report Recommendations

Recommendations from the Initial Usability Evaluation of the APDA Software and Hardware.

General

1. Preventing accidental exiting from the system or resetting of the system clock.
2. Allowing the user easy access to comments.
3. Simplifying the APDA process, e.g., using clear textual descriptors indicating how a user moves from one screen to another.
4. Reducing the need for scrolling, e.g., by providing wrap-around text.
5. Reducing the size of the buttons. This should help reduce the potential for accidental activation and also provide additional space for text display.
6. Using digital ink for comments.
7. Conducting a task analysis. Several of the user's indicated that the information collected via the data collection screen may not be of much value. The simple data collected via the APDA may be a limitation of the size of PDA screens. Conduct a task analysis to try and discover the type of data that the shipboard evaluator needs. If more data, or data of a screen-intense nature, is required then consider investing in a larger handheld or wearable computer device.

Usability Issues Relating to All Screens

A. Heuristic Violation: Minimizing the User' Memory Load

Issue

Text readability/Font size concerns.

Recommendation

1) Use a font size no smaller than 9-point
2) Reduce amount of text displayed in an individual box
3) Increase text box size

Action

1) Add a form that maximizes MOP. Form will be displayed at the touch of a button that is to be part of the "smaller" MOP frame.
2) Increase font size in maximized window for ease of reading.
3) Add an OK button to maximized window to return to Assessment screen.
4) Ensure functionality of the maximized screen and the access button are detailed in training/help materials.

B. Heuristic Violation: Preventing Errors

Issue

When a button/tab is clicked, it turns white (Figure 5 in report)

Recommendation

1) Adding a thick black border around the selected button may make it easier for the user to remember which button is selected. Color could also be used as a highlighting method.

Action

1) Ensure there is a mention that when selected, tabs and buttons will turn white in the training/help material. This is a basic Windows CE function. There is no provision for changing the tab color to something else other than white when selected.

C. Heuristic Violation: Consistency

Issue

Neither the Prebrief nor the Debrief screens have a link to a comment page. However, the user may wish to add a comment while on either of these screens.

Recommendation

1) Adding a comment link to all screens may increase the usability of the software.

Action

1) Add a link. Note: Comments added from the Pre-brief and De-brief pages will not be directly related to a particular MOP as comments added from the Assessment page are. ATEAMS currently has no means for handling comments that are input in either case.

Issue

The titles of the three screens are inconsistently phrased

Recommendation

1) Change the phrasing of the titles of the last two screens to Data Collection and Debrief

Action

1) Rename offending screens.

D. Heuristic Violation: Preventing Errors

Issue

Clicking the Exit button immediately closes the APDA software. No error check message is provided

Recommendation

1) Use the standard Windows error-checking pop-up window

Action

1) Add error-checking functionality.

Issue

The size and location of many of the buttons could lead to errors, i.e., mis-clicks (Figure 6 in report)

Recommendation

1) Reduce the size of the Exit/Next/Previous buttons and insert some space between all buttons.
2) Place the Next/Previous buttons next to each other.
3) Move the Exit/Next and Previous buttons to the extreme lower left/right of the screen.
4) Increase the size of the +/- Event and MOP buttons.
5) Increase space that holds the MOP listing.
6) Make MOP listing area the visual focal point of the user (Consider raising event box higher and enlarging the MOP box).
7) Consider raising the Event box higher and enlarging the MOP box. If possible make the font in the MOP box the largest on the screen so that the user's attention is drawn there.

Action

1) Reduce size of Exit, Next, and Previous buttons. Reposition buttons so that the Exit button is always in the lower right corner of all screens and the Next and Previous buttons are in the lower left of all screens.
2) Increase the size of the +/- Event and MOP buttons
3) Add a form that maximizes MOP. Form will be displayed at the touch of a button that is to be part of the "smaller" MOP frame.

E. Heuristic Violation: Providing Adequate Help and Documentation and Consistency

Issue

APDA provides no access to a help function

Recommendation

1) Add a help function

Action

1) Add a help function

F. Heuristic Violation: Simplicity

Issue

Users want screens to be as simple and efficient as possible

Recommendations

1) Adding a drop-down menu may make it easier for the user to find/select the TE/MOP from the Assessment screen or from the Debrief screen.
2) Comment screen should pop up when as the user clicks the button Other (on the Incomplete screen), since the trainer should indicate what happened to cause a given MOP to be incomplete.
3) Comment screens should be accessible from both the data collection and debriefing screens directly.
4) Users wanted some coding method to be used on the Debriefing screen, indicating where they had made comments related to a given MOP and a means to access those comments directly

Action

1) Add a comment button to the cause code screen to provide the user the option to comment a cause code. Note: there is no ATG requirement to add comments to the cause codes and forcing a trainer to the comment screen each time the cause code "other" is selected will result in unnecessary screen taps. Providing the option to add a comment however is a good idea.
2) Add a comment button to the Pre-brief and Debrief screens
3) Add an indicator that a comment has been entered for a particular MOP (the character C displayed under the MOP in the Assessment and Debrief screens may be a good way to indicate this. Another possibility might be to change the text that appears in the Assessment screen comment button from "Comment" to "Edit Comment" when an MOP that has a comment entered is displayed).
4) Add a Review Comments button to the Debrief screen. The Comment Review screen would display the MOPs commented and the comment text that has been entered for each. The screen would also provide an edit comments button that will allow for editing/addition of comments from the debrief screen.

Issue

On both the Prebrief and Debriefing screens, the user must click on the pluses to open up all the objectives

Recommendation

1) Include 'Expand all' and 'Collapse all' options for objectives

Action

1) eVB does not include "Expand All" and "Collapse All" functionality.

Usability Issues Relating to Prebrief Screen

G. Heuristic Violation: Prevent Errors

Issue

Potential for activating one function over a desired one to crowded positioning of buttons/tabs.

Recommendations

1) Reduce the size of the buttons and insert some space between all buttons

2) Move the Exit/Next and Previous buttons to the extreme lower left/right of the screen

Action

1) Reduce size of Exit, Next, and Previous buttons. Reposition buttons so that the Exit button is always in the lower right corner of all screens and the Next and Previous buttons are in the lower left of all screens.

2) Unable to reduce size of buttons while maintaining readability of button label.

Issue

Abbreviations are inconsistently created

Recommendation

1) Abbreviation method chosen should be consistently applied, with truncation being the preferred method

Action

1) Need fleet feedback to provide labels that will be most meaningful and understandable to end users. Abbreviations and truncations currently used on existing screens are meaningful to training team members however improvements may be gained through fleet feedback.

H. Heuristic Violation: Consistency

Issue

Position of the three aforementioned rows of tabs change, depending on which is clicked

Recommendation

1) Keep the position of each button the same no matter which button is clicked

Action

1) This is standard Windows functionality. No change to the screen will be made. Add a reference regarding the changing tab positions to the training/help material.

I. Heuristic Violation: Speak the User's Language

Issue

Scenario Time displayed for a selected TE is difficult to understand

Recommendation

1) Use the words "Scenario Time" as a descriptor of the displayed time

Action

1) The Scenario Time is currently displayed in a form that is inherently understood by training team members - Start+HH:MM:SS. Add a reference regarding the format of the Scenario Time to training/help material.

<u>Usability Issues Relating to the Collecting Data Screen</u>

J. Heuristic Violation: Consistency

Issue

Relative to the Prebrief and Debrief screens, the size of the Exit button on the Collecting Data screen is inconsistent (larger and conjoined with the Next/Previous buttons)

Recommendation

1) Standardize the size and placement of the common command buttons

Action

1) Standardize the size and placement of the common command buttons (Exit, Next, Previous) between all screens

Issue

There is an inconsistency in the way text is displayed between the Team Events Timeline and Assessment screens. Text in the former wraps. Text in the latter does not.

Recommendation

1) Force the text to wrap to eliminate need for horizontal scroll bar

Action

1) Create a word wrapping function for the Events Timelines.
2) Integrated Events timeline window displays events backwards. This needs to be fixed!

Issue

All caps are used for the text of the TEs when the Assessment button is clicked but when the Timeline button is clicked mixed case is used for the text of the TEs

Recommendation

1) Change the former to mixed case

Action

1) This is not currently a problem. The situation may have been a function of the way the original scenario was developed.

K. Heuristic Violation: Simplicity

Issue

Several users were confused by the textual descriptors of the Next, Previous, + MOP – MOP, + TE, and – TE buttons

Recommendation

1) Next and Previous are frequently used as textual descriptors of navigational aids in web pages as well as in computer based training software. Therefore they should not be changed. Reducing the size of these buttons, while simultaneously increasing the size of the ± MOP and TE buttons may help. Color coding the latter buttons may also help. To capitalize on a users mental model of tape controls, e.g., of a video player, an arrow scheme may also be employed, e.g., $\Leftarrow$ MOP $\Rightarrow$.

Action

1) Reduce size of Exit, Next, Previous buttons and reposition as previously stated.
2) Enlarge size of +/- Event and MOP buttons.
3) Modify label on the +/- Event and MOP buttons to resemble that of a VCR (i.e. <Event, Event>, <MOP, MOP>) to indicate direction of navigation.

Issue

When the user enters the Data Collection screen, the initial screen seen is the Timeline. This seemed to confuse the users.

Recommendation

1) Consider making Assessment the default screen of the Data Collection screen

Action

1) Need fleet feedback to determine most desirable order for presenting screens.

L. Heuristic Violation: Simplicity and Progressive Disclosure

Issue

It should be made obvious to the user how to quickly access the data collection component of the system

Recommendation

1) Label the 'Previous' and 'Next' buttons with the actual screen name that will appear when the button is clicked

Action

1) Changing the button labels contradicts previous recommendation to keep these the same. No change will be made, however comments regarding the order of screen appearance and meaning of navigation buttons will be included in the training/help material when developed.

M. Heuristic Violation: Consistency and Minimize User's Memory Load

Issue

Two of the main issues associated with a PDA are readability and comprehension. The designers of the APDA reduced the font size used in the Assessment screen, relative to the Team Events Timeline screen, presumably to reduce the need to scroll through the text (i.e., so the user can read an entire TE and/or MOP without having to scroll). However, this is inconsistent use of font size and may reduce the readability of the text.

Recommendations

1) The best solution may be to wrap the text and, if necessary, reduce the font size since the user can always move the PDA closer to his eyes

2) Consider raising the Event box higher and enlarging the MOP box

2) If possible make the font in the MOP box the largest on the screen so that the user's attention is drawn there

Action

1) Maximized screen discussed earlier will fix these issues.

Issue

Text box layout used reduces the amount of space available for text, both on the Timeline and the Assessment screen

Recommendations

1) Box can either be eliminated or enlarged such that its borders touch the edge of the PDA screen

2) Control buttons (Team Events, Integrated Events, Previous, etc.) can be reduced in size and moved away from the text box, increasing the amount of available space for the textbox

Action

1) Box border cannot be eliminated. It is part of the basic layout of the screen.

2) Control button issues will be addressed as previously discussed.

Issue

When the Assessment button is clicked, the TE title is duplicated

Recommendation

1) Redundant text can be eliminated, again increasing the amount of available space for the text box

Action

1) The TE name is presented on both Timeline and Assessment pages to reduce the need to cycle between the Timeline screen and Assessment screen. Presenting the TE, TO, EO, and MOP on one page maintains the referential integrity that provides the context needed to evaluate the trainees performance on a particular MOP. This feature will not be changed.

N. Heuristic Violation: Prevent Errors

Issue

The proximity of the Previous, Exit, and Next buttons, relative to the Comment and the ± Event and MOPs buttons, may increase the possibility of mis-clicks

Recommendation

1) Reduce the size of the Previous/Exit/Next buttons and insert some space between all buttons

2) Place the Next/Previous buttons next to each other

3) Move the Exit/Next and Previous buttons to the extreme lower left/right of the screen to help prevent accidental activation

4) Include an error-check for the exit button

Action

1) Add an error check for the Exit button.
2) Previously discussed actions will correct the other issues.

O. Heuristic Violation: Consistency

Issue

Same information is displayed when either the Team Events or Integrated Events button is selected in the Timeline screen. The order of TEs changes depending on which button is clicked.

Recommendation

1) Unless there is a clear purpose for the Integrated Events button, eliminate it. Make sure the order of the events (as displayed to the user) is proper.

Action

1) The evaluated scenario contained only CSTT events. In a case like that, the team timeline and integrated timeline will be identical. The two displays are important for maintaining the overall context of the scenario when more than one training team is participating. No action will be taken.

P. Heuristic Violation: Minimize User's Memory Load

Issue

There is no indication of the total number of TE's or MOPs. There is no indication of the number of TE's/MOP's completed (e.g., percent completed).

Recommendation

1) Attempt to include the aforementioned counters. User should be provided with an indication of the number of complete (or incomplete) TE's and MOPs.

Action

1) Add a counter to indicate TE/MOP completion.

Issue

The position of the two pairs of TE and MOP buttons is mirrored as follows (see Figure 8 above):

- Event – MOP Comment + MOP + Event

This mirroring may confuse the user.

Recommendation

1) A better layout may be as follows:

+ MOP - MOP Comment + Event - Event

An alternative method would be to use toggle buttons, e.g.,

MOP ⊞ ⊞ Event Comment

or a symbol system, e.g.,

⇦ MOP ⇨ ⇦ Event ⇨ Comment

Action

1) Button positioning and Text labels will be modified as previously described.

Issue

Text labeled Event Time is somewhat distant from the selected TE text. Meaning of the text label Event Time may not be clear.

Recommendation

1) Moving the label/event time closer to the TE text may make it easier to find/use.
2) Replace text label of Event Time with the text label of Time of Event - may make the meaning of the TE time a little clearer.

Action

1) Attempt to place the Event Time indicator inside the box that lists TE, TO, EO and MOP.
2) Determine whether Time of Event label will fit in space available.

Q. Heuristic Violation: Consistency and Prevent Errors

Issue

Clock control buttons consist of a reset button and a button that toggles between start and pause. No fast forward, rewind, or stop buttons are provided.

Recommendation

1) May be useful to include rewind, fast forward, and stop buttons.

Action

1) VCR type controls not necessary. An explanation of clock operation will be included in the training/help materials.

Issue

No error check is provided when the reset button is clicked. Thus the user can accidentally click this button and reset the scenario clock to zero.

Recommendation

1) Provide an error check message when the reset button is clicked

Action

1) Provide error check for clock.

R. Heuristic Violation: Simplicity and Prevent Errors

Issue

Method of text input not simple nor intuitive.

Recommendation

1) Use digital ink note text input method.

Action

1) Provide training/help information on text input methods. An ink note capability exists as included functionality w/in the iPaq. Another useful function that has been added is the transcriber that converts handwriting to text.

Issue

Comment screen consists of a blank text box and an OK button. When entering in a comment, the keyboard or the writing surface area covers the OK button of the Comment screen. To reveal the OK button, the user must click either a keyboard or pen icon (located to the immediate left of the aforementioned arrow). If the user accidentally clicks another button, e.g., the demo button found on the handwriting screen, it is quite easy to get lost.

Recommendation

1) Make sure the OK button is always displayed.
2) Rename the OK button to Save.

Action

1) Reposition the OK button.
2) Change the label to Save vice OK

S. Heuristic Violation: Consistency

Issue

No Cancel button available on Comment screen. If no comment is entered, the only option the user has for returning to the Collecting Data screen is to click the OK button. Once the OK button is clicked, it is not known if blank comments are being inserted into the debrief database**.**

Recommendation

1) Add a Cancel button.

Action

1) Add a Cancel button

T. Heuristic Violation: Minimize User's Memory Load

Issue

When a comment has been entered and saved by the user, the user may wish to edit and/or delete comments before data collection has stopped. Once comments are entered there is no way to edit them.

Recommendation

1) Provide a method through which the user can recall and edit notes.

2) The presence and the method for accessing related MOP comments in debrief should be obvious. In other words, provide a symbol near a MOP where a comment has been made. Allow access to that comment by clicking on that symbol.

Action

1) When a comment has been entered, an indicator will appear a discussed above. When the Comment button is selected the software should retrieve the previously entered comment for editing.

<u>Usability Issues Relating to the Debriefing Screen</u>

U. Heuristic Violation: Minimize User's Memory Load

Issue

About 40 characters can fit on the screen. Unfortunately, the menu structure, combined with the length of the MOPs, means that most of the MOP (and sometimes part of the TE) descriptors are not viewable without using the horizontal scroll bar (some MOPs are 19 words long).

Recommendation

1) Based on the TE's and MOPs viewed during this evaluation, it would seem that scrolling horizontally would tax the users memory load to a greater degree than if the text wrapped. However if the length of the text is long, e.g., 19 words, then wrapping the text may also unduly burden the user's memory load. Ideally no scrolling should be used, e.g., a 19 word MOP should fit on the display. For this to occur, however, the font size would have to be reduced, perhaps to an unreadable level. In addition the readability of a given display depends on, among other factors, the amount of stress the user is under, the viewing conditions (light levels, vibration), and quality of the display. Related data would seem to indicate that the use of wrap-around text, drop-down windows, or some combination of the two might be in order (The Windows® Interface Guide, 1995). It is important that some (or all) of the above factors be examined in an operational setting so that the usability of a PDA/pocket PC can be thoroughly evaluated.

Action

1) The Maximized Screen will correct these issues for the MOPs being evaluated. The TE, TO, and EO are merely provided as a fallback in case the trainer cannot remember the

context of the MOP. Since those areas will not be critical to the evaluation of the MOP, they will not be included in the Maximized Screen.

V. Heuristic Violation: Consistency and Simplicity

Issue

There is no template to guide the user for deciphering the MOP code

Recommendation

1) Provide a template to the user (perhaps through a help button).

Action

1) In the Debrief Screen, the software currently provides the first letter of each of the cause codes selected as a reason for the MOP being evaluated as incomplete. This is a good way to concisely indicate which cause codes apply to the MOP and should be readily apparent to any training team member. Add comments explaining the structure of the debrief and the meaning of the incomplete cause codes to the training/help material to avoid any confusion for the user.

Issue

There is no indication if a comment had been entered for a given MOP

Recommendation

1) Add a comment code to the MOP so the user will know which MOP has a comment associated with it.

Action

1) This issue will be resolved as discussed above.

Issue

There is no method for the user to display the comments associated with a given MOP

Recommendation

1) Provide a method through which the user can display his comments.

Action

1) This issue will be resolved as discussed above.

W. Actions Not A Part of the Usability Report

1) Filtering/Displaying Incomplete MOPs

2) Modify Pre-brief to provide the IITSEC specific pre-brief materials.

3) Develop functionality that allows 1 trainer to collect data on more than 1 trainee.

Appendix D: Actions to be taken for HHD – 29 Oct 2001

All actions listed as Priority 1 will be worked together as the top priority. Priority 2 will be worked when we finish the Important to Do Priority 1 actions. We need to move quickly on the Priority 1 items so that we can use the HHD software for the CReaMs test events that are planned for next week.

Important Things to Do

Priority 1

A1 Add a form to the Assessment Screen that maximizes MOP. Form will be displayed at the touch of a button that is to be part of the "smaller" MOP frame.
Addresses: D4, M1, M2, U1

Important - Easy

A2 Increase font size in maximized window for ease of reading.
Addresses: M1, M2

Important - Easy

A3 Add an OK button to maximized window to return to Assessment screen.

Important - Easy

D1 Add error-checking functionality when Exit button is pushed and when resetting clock.
Addresses: N1, Q2

Important - Easy

J3 Integrated Events timeline window displays events backwards. This needs to be fixed!

Important – Easy/Medium

Priority 2

C1 1) Add a comment link to prebrief and debrief screens. 2) Store Comments. Note: Comments added from the Pre-brief and De-brief pages will not be directly related to a particular MOP as comments added from the Assessment page are. 3) ATEAMS currently has no means for handling comments that are input in either case.
Addresses: F2

1) Important – Easy
2) Important – Medium
3) Important – Impossible for now

F1 Add a comment button to the cause code screen to provide the user the option to comment a cause code. Note: there is no ATG requirement to add comments to the cause codes and forcing a trainer to the comment screen each time the cause code "other" is selected will result in unnecessary screen taps. Providing the option to add a comment however is a good idea.

Important - Easy

F4 1) Add a Review Comments button to the Debrief screen. The Comment Review screen would display the MOPs commented and the comment text that has been entered for each. 2) The screen would also provide an edit comments button that will allow for editing/addition of comments from the debrief screen.
Addresses: V3

1) Important - Medium
2) Nice to have - Hard

Priority 3

R2 Reposition the OK button on the Comment Screen to remain clear of pop-up keyboard at all times.

Important - Easy

W1 Filtering/Displaying Incomplete MOPs

Important - Medium

<u>Nice to Have Things to Do</u>

Priority 1

D2 Reduce size of Exit, Next, and Previous buttons. Reposition buttons so that the Exit button is always in the lower right corner of all screens and the Next and Previous buttons are in the lower left of all screens.
Addresses: G1, G2, J1, K1, M3, N2, P2

Nice to have - Easy

D3 Increase the size of the +/- Event and MOP buttons.
Addresses: K2, P2

Nice to have - Medium

K3 Modify label on the +/- Event and MOP buttons to resemble that of a VCR (i.e. <Event, Event>, <MOP, MOP>) to indicate direction of navigation.
Addresses: P2

Nice to have - Easy

Priority 2

F3 Add an indicator that a comment has been entered for a particular MOP (change the text that appears in the Assessment screen comment button from "Comment" to "Edit Comment" when displaying an MOP that has had a comment previously entered. If the "Edit Comments button is selected the button will retrieve the previously entered comment).
Addresses: T1, V2, V3

Nice to have - Medium

F4 1) Add a Review Comments button to the Debrief screen. The Comment Review screen would display the MOPs commented and the comment text that has been entered for each. 2) The screen would also provide an edit comments button that will allow for editing/addition of comments from the debrief screen.
Addresses: V3

1) Important - Medium
2) Nice to have - Hard

Priority 3

None.

Priority 4

C2 Rename inconsistently named screens.

Nice to have - Easy

J2 Create a word wrapping function for the Events Timelines.

Nice to have – Easy/Medium

P1 Add a counter to indicate progress in TE/MOP completion.

Nice to have - Hard

P3 Attempt to place the Event Time indicator inside the box that lists TE, TO, EO and MOP.

Nice to Have - Medium

R3 Change the label on the Comment Screen OK button to Save vice OK.

Nice to have - Easy

S1 Add a Cancel button to the Comment Screen.

Nice to have – Easy/Medium

W2 Modify Pre-brief to provide the IITSEC specific pre-brief materials.

Nice to have - Medium

W3 Develop functionality that allows 1 trainer to collect data on more than 1 trainee.

Nice to have - Hard

<u>Not Important Things to Do</u>

Priority 4

P4 Determine whether Time of Event label will fit in space available.

Not Important - Easy

Create a Help/Training Function – Important for Integrated. Will not be done for I/ITSEC.

A4 Ensure functionality of the maximized screen and the access button are detailed in training/help materials.

B1 Ensure there is a mention that when selected, tabs and buttons will turn white in the training/help material. This is a basic Windows CE function. There is no provision for changing the tab color to something else other than white when selected.

E1 Add a help function.

H1 Positioning a row of tabs at the front of the grouping when selected is standard Windows functionality. No change to the screen will be made. Add a reference regarding the changing tab positions to the training/help material.

I1 The Scenario Time is currently displayed in a form that is inherently understood by training team members - Start+HH:MM:SS. Add a reference regarding the format of the Scenario Time to training/help material.

L1 Changing the button labels contradicts previous recommendation to keep these the same. No change will be made, however comments regarding the order of screen appearance and meaning of navigation buttons will be included in the training/help material when developed.

Q1 Changing button labels on the clock to VCR type controls not necessary. An explanation of clock operation will be included in the training/help materials.

R1 Provide training/help information on text input methods. An ink note capability exists as included functionality w/in the iPaq. Another useful function that has been added is the transcriber that converts handwriting to text.

V1 In the Debrief Screen, the software currently provides the first letter of each of the cause codes selected as a reason for the MOP being evaluated as incomplete. This is a good way to concisely indicate which cause codes apply to the MOP and should be readily apparent to any training team member. Add comments explaining the structure of the debrief and the meaning of the incomplete cause codes to the training/help material to avoid any confusion for the user.

No Action to be Taken

F5 eVB does not include “Expand All” and “Collapse All” functionality for treelists.

G3 Need fleet feedback to provide labels that will be most meaningful and understandable to end users. Abbreviations and truncations currently used on existing screens are meaningful to training team members however improvements may be gained through fleet feedback.

J4 No action.

K4 Need fleet feedback to determine most desirable order for presenting screens when Assessment screen is initially called up.

M4 No action.

O1 No Action.

References

Data to go. (2001, May). *Consumer Reports*, 20-23.

Eberts, R. E. (1994). *User interface design*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Gravlee, C. C. (2002). Mobile computer-assisted personal interviewing with handheld computers: The Entryware System 3.0. *Field Methods* 14(3), 322-336.

Hamel, C. J. & Clark, S. L. (1986). CAI evaluation checklist: Human factors guidelines for the design of computer-aided instruction. *Technical Report: NAVTRASYSCEN TR86-002*. Navy Training Systems Center – Human Factors Division: Orlando, FL.

Lynch, P. J., & Horton, S. (1999). *Web style guide: Basic design principles for creating web sites*. New Haven, CT: Yale University Press.

Lyons, D. M., & Allen, R. C. (2000). Mobile aid for training and evaluation (MATE): A hand-held, configurable set of team performance measurement tools. *Proceedings of the 2000 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*, (pp. 661-671).

Mandel, T. (1997). *The elements of user interface design*. New York: John Wiley & Sons, Inc.

Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.

Noldus Information Technology (2003, July 10). *Noldus products*. Retrieved July, 10, 2003, from http://www.noldus.com/products/index.html

RCI (Producer). (2000). Battle force tactical training (BFTT) system overview [CD ROM]. Virginia Beach, VA: Combat Direction Systems Activity, Dam Neck Naval Surface Warfare Center.

Ricci, K., Allen, R.C., Reynolds, A., Daskarolis-Kring, E., & Hodak, J. (2001). *An evaluation of the AFLOAT training exercise and management system (ATEAMS) software*. Orlando, FL: Naval Air Warfare Center Training Systems Division.

Sanders, M. S. & McCormick, E. J. (1987). *Human factors in engineering and design*. New York: McGraw-Hill, Inc.

Stretton, M. (2001). Afloat training, exercise, and management system(ATEAMS): Enabling objective-based training. *Proceedings of the 2001 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC),* (pp.1429-1519).

The Windows®Iinterface Guidelines for Software Design. (1995). Redmond, WA: Microsoft Press.

Weber, B. A., Roberts, B. L. (2000, May/June). Data collection using handheld computers. *Nursing Research, 49*(3), 173-175.

Waterson, S., Matthews, T., & Landay, J. A. (2002). In the lab and out in the wild: Remote web usability testing for mobile devices. *Extended Abstracts of Human Factors in Computing Systems: CHI 2002*, Minneapolis, MN, April 20-25.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins Publishers, Inc.

Acronym List

APDA – ATEAMS Personal Digital Assistant
ATEAMS – Afloat Training Exercise and Management System
BFTT – Battle Force Tactical Trainer
EO – Enabling Objective
GUI – Graphical User Interface
MOP – Measure of Performance
OBT – Objective Based Training
PDA – Personal Digital Assistant
ROE – Rules of Engagement
TE – Training Event
TO – Terminal Objective

Supporting Observers During Distributed Team Training - The Development Of A Mobile Evaluation System

Marcel P.W. Van Berlo, MSc
Annemarie M.F. Hiemstra, MSc
Ing. Wytze Hoekstra
TNO Human Factors, Department of Training and Instruction
PO box 23
3769 ZG Soesterberg
The Netherlands
E-mail: vanberlo@tm.tno.nl

Summary

Distributed team training, often in joint settings, is becoming more and more important in the military training today. As the teams and training staff are not physically on the same location, special attention should be paid to performance measurement and feedback. The question is what should be measured in distributed training and how the training staff should be supported in doing this. This paper describes a mobile performance measurement and evaluation tool, specifically developed for distributed team training (MOPED). The MOPED tool helps the evaluator in observing team performance and in quickly generating, sending and receiving data to support his evaluation.

Introduction

Distributed team training, often in joint settings, is becoming more and more important in military training today. As the subteams and team members are not physically on the same location, performance measurement and providing feedback could be problematic. In order to give adequate feedback, it is essential that observers, who are distributed themselves as well, can quickly compare and integrate their observations. In this way, the time needed to prepare the after action review can be reduced to a minimum. The sooner the results of an exercise can be evaluated, the better it is. In a distributed context, there should be a relation between the performance of the own team and the higher level team in order to evaluate and improve the co-ordination and teamwork. Although systems for automated team performance measurement are becoming more available, it is still the human observer that is responsible for measuring team processes. However, measuring team processes is a difficult task. Besides, the risk of having several observers is multiple interpretations of the same observed behavior. Therefore, the issue is what should be measured in distributed training and how the training staff should be supported in doing this. This paper describes a mobile performance measurement and evaluation tool, specifically developed for distributed team training (Hiemstra, Van Berlo & Hoekstra, 2003). The tool helps the evaluator in observing team performance and in quickly generating, sending and receiving data to support the evaluation. In the next section, the characteristics of distributed teams will be described, as well as the performance measurement. Next, the methods and results of specifying the mobile tool will be illustrated. This is followed by a brief discussion about the first try-out of MOPED. Finally, some concluding remarks will be made.

1. Characteristics of distributed teams

A distributed team consists of several subteams and/or individuals working together on a joint mission, but who are not able to meet face-to-face because of differences in space and/or time. Distributed teams are characterized by the fact that they are geographically distributed, electronically linked, and functionally

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held in Genoa, Italy, 13 – 15 October 2003, and published in RTO-MP-HFM-101.

and/or culturally divers (Schraagen, 2001). In many respects, distributed teams are not different from teams conducting their tasks in a face-to-face environment. In both cases, the teams are serving the organization in achieving a certain, common, goal. However, installing distributed teams makes certain aspects of team performance more explicitly visible: who is responsible for what, how is the balance between self-direction and central management, how can teamwork be optimized between all subteams and individuals. Preparation before task performance as well as a common understanding on the various roles (convergence) seem to be crucial factors in the success of distributed teams (Rocco, 1998). The same holds true for the use of technological tools that links the team members. The diversity of teams can have a significant impact on the team's performance. Especially in the case of joint missions, involving teams with different organizational or national backgrounds, cultural differences can cause many problems. Multicultural teams should therefore explicitly formulate group norms and identify how these cultural differences could affect the cohesiveness and effectiveness of the team (Van Vliet, 2003).

Leadership in a distributed environment is critical in both the planning and the execution phase of the mission (Cook & Klumper, 1999). The emphasis should be on relationships within the distributed team, or with other words, on communication and interpersonal skills. The leader/commander should be open for advise, and he should have the competency to adopt other team members' perspectives (Cook & Klumper, 1999). This is difficult in a distributed environment because the context is incomplete. Team leaders should therefore explicitly indicate what issues they want to be informed about, when they want to be involved, and to what extent team members can make decisions on their own (Van Bree, Vlietman & Wierda, 2001).

A team can be effective on two levels: the task and the process level. On the task level, performance assessment is relatively easy because the task result can be measured objectively (e.g. completion of a mission, operating the equipment). This is more difficult for processes within a team or between teams. Previous research has shown that communication, information exchange, initiative/leadership and supporting behavior (Smith-Jentsch *et al.*, 1998) are important dimensions of team performance on the process level. Besides, Orasanu (1990) showed that better performing teams are better planning their actions. These aspects seem even more important for distributed teams, due to the lack of face-to-face contact. Singer et al (2001) investigated the differences between local and distributed teams. They found that in all cases the distributed teams did not outperform the local teams. Their explanation for this result was the lack of informal, and face-to-face, communication between the missions. Finally, Singer et al. (2001) stress the relevance of communication: because of the geographical distribution of team members, the interdependency increases, and therefore the need to communicate in order to achieve the shared and common goals.

2. Team performance measurement and feedback

To improve the training effectiveness, performance measurement and feedback should be conducted systematically and in a standardized manner across all distributed teams (Dwyer *et al.*, 1997; Fowlkes *et al.*, 1994). As the teams and training staff are not physically on the same location, special attention should be paid to performance measurement and feedback. Observations of the processes within the distributed team should be consistent among the various (distributed) observers in order to provide consistent feedback to the team members. The question is what should be measured in distributed training and how the training staff should be supported in doing this? Previous research has shown that communication, information exchange, team leadership and supporting behavior (Smith-Jentsch et al., 1998) are important dimensions of team performance. These dimensions are even more important for distributed teams, due to the lack of face-to-face contact. In addition, preparation before task performance as well as a common understanding of the various roles (convergence) seem to be crucial factors in the success of distributed teams (Rocco, 1998).

Previous research on team performance measurement resulted in the Command and Control Process Measurement Tool (C2PMT; Van Berlo & Schraagen, 2000; Schraagen & Van Berlo, 2001). The C2PMT is a generic checklist comprising standards a command & control team should meet. Every standard is briefly clarified and explained in order to ensure a uniform interpretation by the evaluators: it describes the contents and coverage of the standard and, if applicable, the relation with other standards. For every standard, performance indicators have been formulated giving concrete form to the standard enabling the evaluators to observe and interpret the team processes. These performance indicators are formulated concisely, and are easily scored in terms of whether the behavior was observed or not. The evaluator can explain and illustrate every observation: this contains both positive and negative examples being observed. Inclusion of these example behaviors is important for providing feedback in the final written report and for enhancing learning opportunities. The C2PMT was adjusted for distributed team training (C2PMT-Distributed). With this tool the evaluator can score targeted behaviors that are both important within the team and between the distributed teams. Based on previous research on the Mobile Aid for Training and Evaluation (Pruitt et al., 1997; Lyons & Allen, 2000), the C2PMT-D was implemented on a hand-held device. The functional specifications of this MOPED tool (MObile Performance measurement and Evaluation of Distributed team training) will be discussed in the next section.

3. Functional specifications of the MOPED tool

The MOPED tool comprises several generic categories on which the performance of a distributed team can be assessed. Every category is given concrete form by specific performance indicators. Completeness and usability of the performance indicators are balanced as well as possible. If possible, redundancies are avoided in order to prevent the same behavior being rated in several categories. In order to offer a tool that can easily be managed by observers, the total number of categories has been kept to a minimum. The categories are described below, illustrated with some performance indicators.

a) Preparation: availability and quality of plans and checklists, concurrence of the plans and checklists between all distributed teams.
b) Information flow within a team: modality and timing of information exchange.
c) Information flow between teams: modality and timing of information exchange, informing the right persons/teams.
d) Use of ICT systems: adequately using the communication devices, attunement of various ICT systems.
e) Team decision making: indicating expectations by overall commander, stimulating participation by asking suggestions, clear and unambiguous decisions, stating the right priorities.
f) Active monitoring of critical tasks: check deadlines, asking for clarifications.
g) Interpersonal relations and mutual support: every team is actively involved, good atmosphere within the team, supporting each other without neglecting one's own task.
h) Back up facilities: appointment of alternate commander, no interruption of the chain of command, back up of teams.

The MOPED-tool should not only facilitate the performance measurement, but the debriefing as well. In this respect it is important to know how the quality of the own team's performance relates to the performance of the other teams within the distributed context. However, especially during large-scale exercises with teams physically dispersed over a large area, the communication and information exchange between the also distributed observers is difficult. Therefore, the MOPED tool should facilitate the mutual data exchange between observers, and be able to receive and present these data in a format to be used during the debrief. Based on these demands, a pocket-PC has been chosen rather than a PDA or tablet-PC. The display of a PDA was considered to be too small, while a tablet-PC was regarded as too large to handle during field exercises. A pocket-PC can be easily carried, the display is large enough to get a clear overview, and a keyboard facilitates making notes/explanations. Figure 1 shows a picture of the MOPED tool. In the remainder of this section, the functionality of the MOPED tool is described. A more detailed discussion can be found elsewhere (Hiemstra, Van Berlo & Hoekstra, 2003).

Figure 1: The MOPED tool

The MOPED tool consists of six parts, displayed as buttons on the opening screen (see Figure 2). Selecting a button by tapping the stylus on the touch screen, activates the respective part of the program:

- New: the observer logs in, and indicates that a new session will start.
- Questionnaire: the checklist with performance indicators.
- Grading: overview of gradings.
- AAR: overview of observations indicated as relevant for the after action review.
- Send data: transmitting data to a central database.
- Results: observers' results of other teams, received ant integrated by central database.

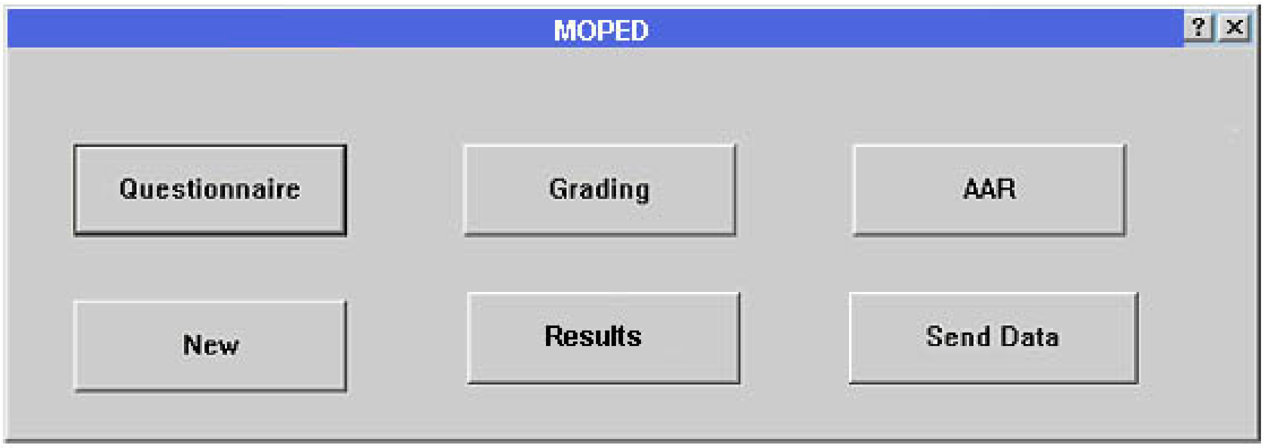


Figure 2: Opening screen of the MOPED tool

Figure 3 shows a screen dump of the C2PMT-D questionnaire. On the left-hand side, the buttons with numbers 1-8 direct the observer to the eight categories the C2PMT-D is comprised of. The category that is selected is depicted on the display. Every performance indicator can be checked as either observed (Yes) or not observed (No). In case the observer wants to include a performance indicator in the AAR, he checkmarks the AAR-button. On the right-hand side, the observer explains and illustrates the rating: this contains both positive and negative examples being observed. Tapping on an explanation-box presents a pop-up window enabling the observer to type in more elaborate text. Inclusion of these example behaviors is important for providing feedback in the final written report and for enhancing learning opportunities. Completely to the right are two arrows enabling the observer to scroll up and down through all performance indicators. After having filled out all performance indicators, an overall grading can be determined for the respective category (EX: excellent, ST: satisfactory, MA: marginal, UN: unsatisfactory, NG: not graded). During the course of an exercise, the observer can easily switch between all categories by tapping on the eight numbers. Tapping on the Ready-button leads the observer back to the opening screen.

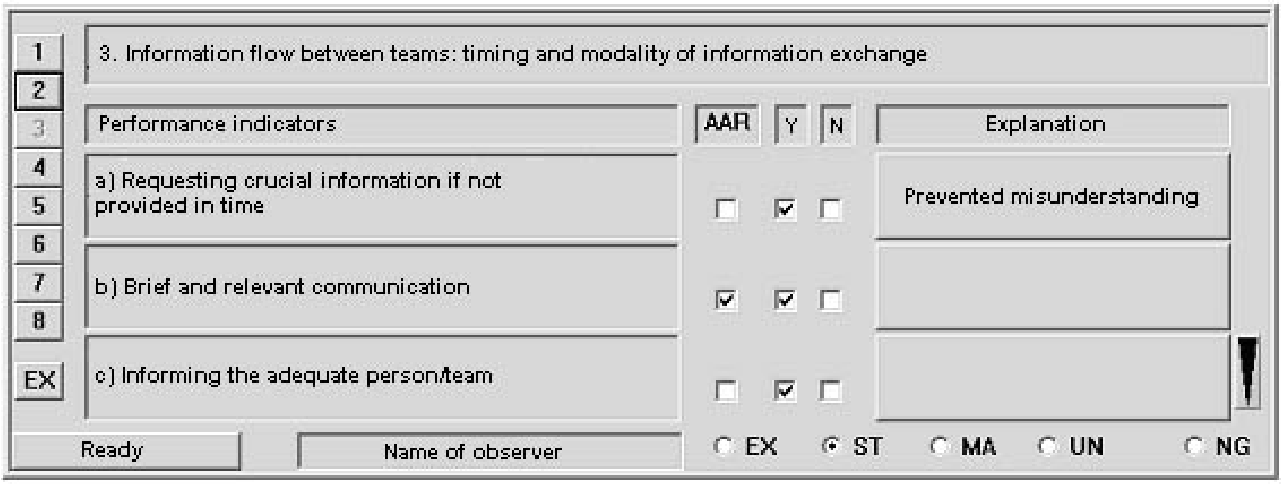


Figure 3: The C2PMT-D questionnaire

All checkmarks and explanations made by the observer are automatically linked to the other parts of MOPED. The overall gradings are linked to the 'Grading' overview. This overview depicts every category and its overall grading, providing for a concise summary of the team's performance (see Figure 4).

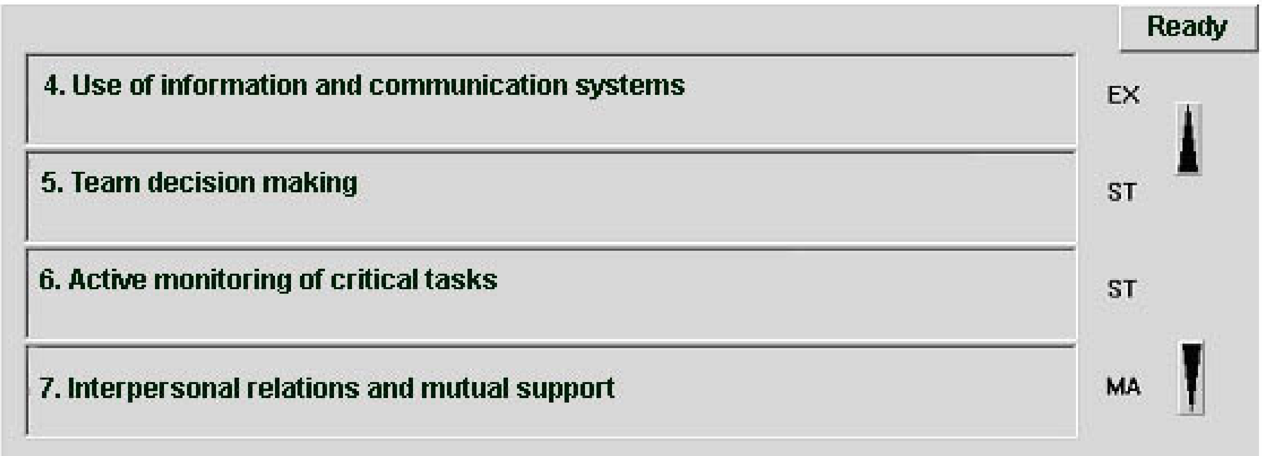


Figure 4: The Grading overview

After an exercise, the observer can send his observational data to the central database ('Send data'). In this central database, all observational data of all distributed observers are gathered and integrated. Only results that are explicitly related to the performance of the distributed team and not a local team (inter vs. intra) are sent back to the observers in the field. Figure 5 shows a screen dump of the 'Results'. The numbers behind every performance indicator show how many observers scored 'Yes' or 'No'. The box behind it show the remarks and explanations made by every other observer. Tapping on it will enlarge this box.

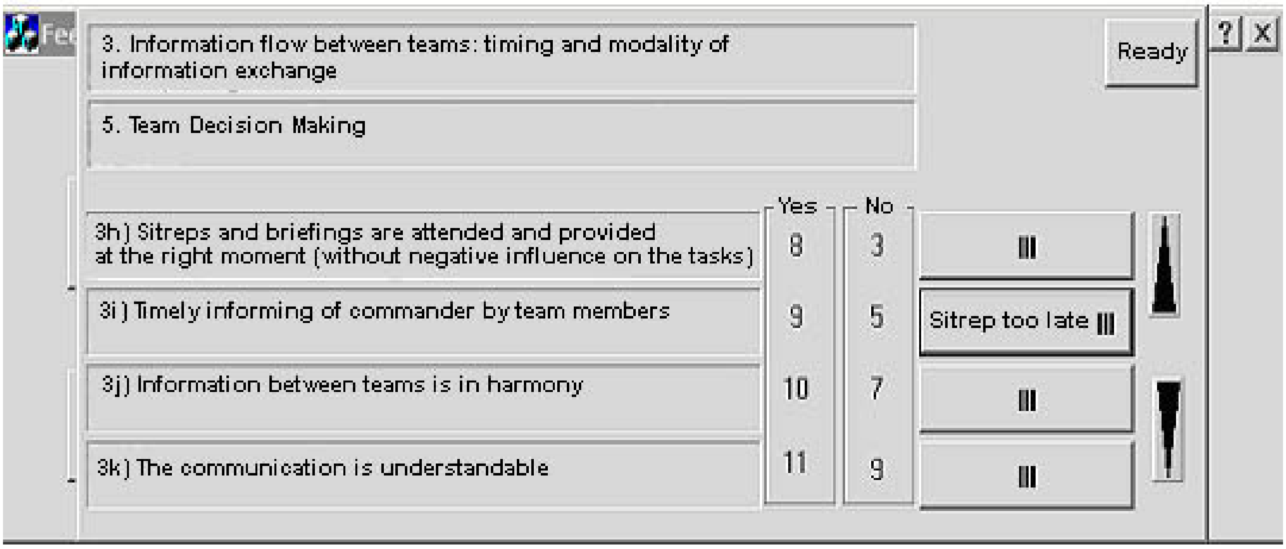


Figure 5: The Results overview

In case the observer wants to include a performance indicator in the after action review, he checkmarks the AAR-button behind it. An overview of all these performance indicators is provided in the AAR-part of MOPED. This overview helps the observer in preparing and conducting the after action review. Figure 6 shows a screen dump of the AAR-overview.

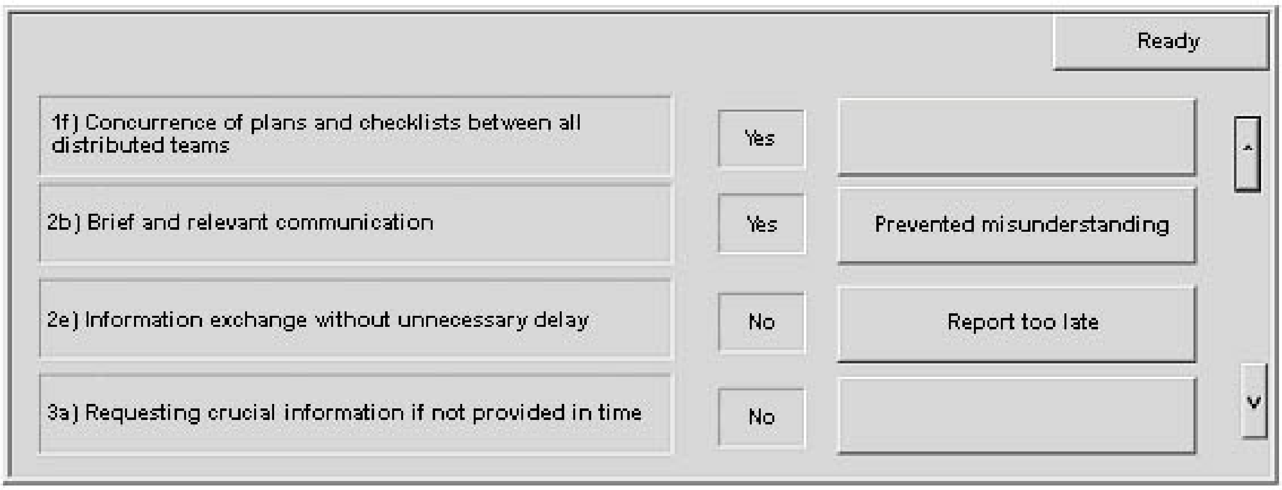


Figure 6: The AAR overview

5. Try-out

A try-out has been conducted at the Operational School of the Royal Netherlands Navy for a period of two weeks (for a more detailed discussion, see Van Rijk, Hiemstra & Hoekstra, 2003). The team was the command central team responsible for the defense of a frigate. This team is divided into two teams who have to work together, but have their own tasks as well: one covering the Anti Air Warfare domain, and the other the domain of Anti Surface/Subsurface Warfare. Although the teams share the same command central on board, during off shore training exercises they use two separate simulators. Simultaneously training the distributed teams using a shared virtual environment and the same training scenario provided the opportunity to conduct a small scale try-out of the MOPED-tool as well as the C2PMT-D questionnaire. For this purpose, the generic version of the C2PMT-D has been adapted to this specific domain and this specific Navy command team. Further, a less dichotomous way of scoring has been implemented: in the try-out version, performance indicators can be scored on various scales rather than only ‘yes’ or ‘no’. Besides on the pocket-PC, MOPED was also be installed on a tablet-PC. In this way, the advantages and disadvantages of the man machine interface and the usability of the hand held computers could be evaluated. Four trainers tested MOPED during 2 weeks in 32 training sessions. Data were collected with a questionnaire, observations and interviews.

It was the first try-out of MOPED, so it was inevitable to have several disadvantages. The most important one was that the participants suggested improving the debrief facility. MOPED offers a fixed sequence of items based on the ranking in the questionnaire. The participants, however, wanted to sequence and cluster the items themselves. With respect to the Pocket PC, it was difficult to keep an overview. A disadvantage of the Tablet PC was the weight and the heating up after a while.
One of the advantages was that MOPED enables a quick data exchange between the observers. Further, the team's result was immediately computed, and MOPED supported a quick arrangement of items for the debriefing. Further, the tool was easy to use, with no flipping around of paper pages. With respect to the C2PMT-D items, the results were not clearly interpretable: most items were perceived as valuable, but during several training sessions the performance measurement appeared to focus only on the commanders of the two teams, meaning that not all items of our questionnaire were actually filled out.

6. Conclusion

The C2PMT-D is a generic method that can improve the quality of the observations made during distributed mission training because of the standardized format in which targeted behaviors are scored. Implementation of this method in the MOPED tool supports the observers in quickly generating and processing performance data and organizing the after action review. Based on the experiences with the C2PMT (Van Berlo & Schraagen, 2000; Van den Bosch & Van Berlo, 2002), applying the MOPED tool will have the following advantages:

- A clear insight into the command and control process of distributed teams.
- An objective assessment of the distributed team's command and control process based on the performance indicators.
- Support of relatively inexperienced observers.
- The results of an exercise are easier to interpret.
- The results of various exercises are mutually comparable.
- The lessons learned can be determined more easily.
- Follow-on actions can be determined in a more structured way.

The MOPED tool provides the means for gathering facts and data concerning the quality of command and control processes, and therefore has the potential to determine follow-on actions not exclusively related to training issues, but also to the real-life (quality) management of organizational processes. Possible follow-on actions could relate to, for instance, the quality and availability of checklists, the attunement of various ICT systems, and the security awareness of the personnel.

7. Further Research

Further research will be conducted on the development and validation of the various parts of the MOPED tool. At the moment, the MOPED tool is a prototype with a generic version of the C2PMT-D. Try-outs during several exercises, combined with experiments, are required to improve the quality and usability of both the method and the tool. This will bring us to a better understanding of performance measurement and feedback in distributed team training, and a validation of the method.

Further, we want to improve the support for the AAR. In the current version of the MOPED tool, the organization of the after action review is supported by the 'Grading' and the 'Results' overviews, showing the results of the team's performance in isolation and in the context of the distributed team. This covers the content of the AAR, but not the process of the debriefing itself. Further research will be conducted on methods to facilitate and guide an after action review of a distributed team, and how these methods could be supported by the MOPED tool.

Finally, integrating the MOPED tool into existing technological frameworks is another possibility. One such framework is MIND, developed by the Swedish Defence Research Agency (FOI). The MIND visualization framework has been developed based on the need to handle large amounts of data pertaining to diverse aspects of distributed rescue operations (Morin, 2002). MIND uses data collected in the real environment, regardless if it is a live operation or a training exercise, to construct a time-synchronized, discrete-event representation of the course of events of the operation. The resulting model is a multimedia representation of the distributed tactical operation that can be presented in a visualization tool that supports time-based navigation and animation using multiple views. MIND has been used for supporting after-action reviews in different settings (Crissey, Morin & Jenvald, 2001; Thorstensson et al., 2001). TNO Human Factors and FOI are currently in the process of combining their expertise of human factors knowledge, computer science and training experience. Further research could aim at the integration of the MIND system and the MOPED tool in order to enhance the quality and speed of the data collection and analysis (Van Berlo et al., 2003). This should result in achieving the optimal combination of automated performance measurement and feedback and human observations.

References

Cook, M.J., & Klumper, W. (1999). Metacognitive, social and interpersonal skills and aptitudes in officer performance with distributed teams. *Proceedings of the RTO Human Factors Medicine Panel Workshop 'Officer Selection', November 1999.*

Crissey, M.J., Morin, M., & Jenvald, J. (2001). Computer-Supported Emergency Response Training: Observations from a Field Exercise. *Proceedings of the 12th International Training and Education Conference,* ITEC 2001, Lille, France.

Dwyer, D.J., Fowlkes, J.E., Oser, R.L., Salas, E., & Lane, N.E. (1997). Team performance measurements in distributed environments: the TARGETs methodology. In: M.T. Brannick, E. Salas, C. Prince, (eds.), *Team performance assessment and measurement- theory, methods and applications* (Chapter 7). London: Lawrence Erlbaum Associates.

Fowlkes, J.E., Lane, N.E., Salas, E., Franz, T., & Oser, R. (1994). Improving the measurement of team performance: the TARGETs Methodology. *Military Psychology, 6 (1), 47-61.*

Hiemstra, A.M.F., Van Berlo, M.P.W., & Hoekstra, W. (2003). *Performance measurement and feedback during distributed team training* (in Dutch) (Report TM-03-B006). Soesterberg, the Netherlands: TNO Human Factors.

Van Rijk, R., Hiemstra, A.M.F., & Hoekstra, W. (2003). *Development and try-out of MOPED, a method and tool for performance measurement and feedback in distributed teams* (in Dutch) (Draft report). Soesterberg, the Netherlands: TNO Human Factors.

Lyons, D.M., Allen, R.C. (2000). Mobile aid for training and evaluation (MATE): a hand-held, configurable set of team performance and measurement tools. *Proceedings of the 2000 Interservice/Industry Training, Simulation and Education Conference*, pp.661-671.

Morin, M. (2002). *Multimedia representations of distributed tactical operations.* Linköping studies in science and technology, Dissertation No. 771, Linköping: Linköping University.

Orasanu, J.M. (1990). *Shared mental models and crew decision-making* (CSL Tech. Report No. 46). Princeton, NJ: Princeton University, Cognitive Science Laboratory.

Pruitt, J.S., Burns, J.J., Wetteland, C.R., & Dumestre, T.L. (1997). SHIPMATE: Shipboard Mobile Aid for Training and Evaluation. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting – 1997.*

Rocco, E. (1998). Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. *Proceedings Computer-Human Interaction [CHI]*, Los Angeles, CA, April 1998 (pp. 492-502).

Schraagen, J.M.C. (2001). *Teams at a distance* (Internal document). Soesterberg, the Netherlands: TNO Human Factors.

Schraagen, J.M.C, & Van Berlo, M.P.W. (2001). C2 Assessment tool for large-scale Air Force exercises. P*roceedings of the 6th International Command and Control Research and Technology Symposium.* U.S. Naval Academy, Annapolis, Maryland, June 19-21, 2001

Singer, M.J., Grant, S., Commarford, P.M., Kring, J.P., & Zavod, M. (2001). *Team performance in distributed virtual environments* (Technical report 1118). Virginia: U.S. Army Research Institute for the Behavioral and Social Sciences.

Smith-Jentsch, K.A., Johnston, J.H., & Payne, S.C. (1998). Measuring Team-Related Expertise in Complex Environments. In: J.A. Cannon-Bowers & E. Salas (1998) (Eds.). *Decision Making under Stress: Implications for Individual and Team Training.* Washington, DC: American Psychological Association.

Thorstensson, M., Tingland, B., Björneberg, A., & Tirmén-Carelius, M. (2001). Visualization of an Interagency Exercise in the Stockholm Underground. *Proceedings of The International Emergency Management Society's Eighth Annual Conference*, TIEMS 2001. June 19-22, Oslo, Norway.

Van Berlo, M.P.W., & Schraagen, J.M.C. (2000). A generic assessment tool for evaluating C2 exercises. *Proceedings of the 22nd Interservice/Industry Training Systems and Education Conference (I/ITSEC),* Orlando, Florida, 28-30 November 2000 (pp. 652-660).

Van Berlo, M.P.W., Thorstensson, M., Schaafstal, A.M., Morin, M., Schraagen, J.M.C., & Jenvald, J. (2003). Improving learning from emergency management training: Sweden and the Netherlands are teaming up. *Proceedings of the International Conference on Emergency Management Towards Co-operation and Global Harmonization (TIEMS 2003), pp 278-288*. June 3-6, 2003, Sophia Antipolis, France.

Van Bree, J., Vlietman, M., & Wierda, R. (2001). *Reality in Virtual Teams* (in Dutch). Amersfoort, the Netherlands: Twynstra Work Innovation.

Van Vliet, A.J. (2003). *The effectiveness of AWACS multi-cultural teams* (internal document). Soesterberg, the Netherlands: TNO Human Factors.

Technology Insertion Process: Determining Task Deficiencies and Matching to Effective Technologies

Capt Gregory J. Sidor, David R. Greschke, LtCol Terence S. Andre, Capt Ira A. Schurig, 1Lt Jonathan M. Cain, Maj Justine N. Good, Dr. Winston Bennett, Jr.
Air Force Research Laboratory
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212
(480) 988-6561
fax: (480) 988-6285
email: firstname.lastname@williams.af.mil

Summary

No other time in history have we seen as much readily available technology for training warfighters as we see today. There is an abundance of training technology in almost every warfighter domain. However, the challenge still remains as to how best to use this technology, and in particular, what tasks can benefit the most from insertion of technology into the training environment. The Air Force Research Laboratory (AFRL) in Mesa, Arizona recently made substantial progress toward this challenge by identifying how simulation technologies could be used to enhance flying within the U.S. Air Force, Air Education and Training Command (AETC). Researchers from AFRL developed a technology insertion process that included instructor pilot workshops, visits to flying training bases, training task definitions, assessment of current simulators, surveys to assess critical training tasks, and a Quality Function Deployment (QFD) to prioritize training technologies. The workshops were a critical component of the technology insertion process as they helped us identify the training requirements and tasks for several aircraft training courses including the T-6, T-38, T-1A, AT-38B, T-38C, F-15, and F-16. After the instructor pilots validated the task requirements, AFRL launched an internet-based survey to over 700 instructor pilots across the U.S. Air Force. The survey was designed to collect ratings on task difficulty, syllabus time allocation, how often tasks contribute to busted check rides, proficiency of graduates, and the adequacy of current simulation devices. This data identified the more critical training needs and formed the basis of an index for weighting the criteria in the subsequent QFD. QFD workshop participants included former and current instructor pilots, researchers, and engineers who have had exposure to updated training technologies. Participants in the QFD workshop assessed how useful 27 different advanced simulation technologies would be for training each task. The priority weights from the survey were then used as multipliers for each technology score to yield total weighted scores for each technology and a prioritized rank ordering of those technologies. This analytic process yielded valuable information to aid leaders in making decisions about technology investment. In this paper, we focus on the technology insertion process and how this process can be used to prioritize advanced simulation technology in a flying training environment.

1.0 Backround

In the last few years, rapid advances in simulation technology have provided more capability, especially in the areas of visual systems, networking, and realistic databases. However, a great deal of the current simulation training technology in the United States Air Force (USAF) has not been updated to take advantage of efficiencies offered by some of these newer technologies.

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

In November 2001, the USAF Air Education and Training Command (AETC) asked the Air Force Research Laboratory (AFRL) to assess the current state of flying training in AETC and to identify and analyze opportunities where advanced simulation technologies could be used to increase the efficiency and effectiveness of AETC flying training programs. The Command's motivation for the study stemmed from a desire to build a simulation technology integration roadmap that will guide the development and enhancement of flying training. The results of the study will assist modernization planners in the development of a technology integration roadmap that will identify the plan for insertion of those technologies with the greatest potential to assist in the capability to train, rehearse, refine, and remediate tasks in a safe, effective instructional simulation environment.

Systematically gathering data on training needs allowed for the prioritization of advanced training technologies to be implemented across AETC. This effort provided a first look at potential targets of modern and high fidelity training technology insertion in AETC Specialized Undergraduate Pilot Training (SUPT), Introduction to Fighter Fundamentals (IFF), and fighter Flying Training Units (FTUs). These training system enhancements could help AETC by:

- Reducing the number of busted check-rides, student wash-backs, and student wash-outs
- Training students to a higher level of proficiency than current course outputs
- Ultimately producing pilots that are more combat mission ready due to a higher level of mastery across a greater number of skills, tasks, and competencies

This study provided quantitative and qualitative survey data from the targeted IP populations, onsite evaluations of all ground-based simulator systems flown in AETC, and technology utility assessment scores from current instructor pilots. Scientists, engineers, and subject-matter experts from AFRL have made recommendations on technology applications based on analysis of the data and knowledge of the current state of the art for the various technologies being considered.

2.0 Evaluations of Current Simulators

One of the critical aspects of this study was an evaluation of the current simulation technologies being used at operational training sites. Scientists and engineers from AFRL visited six different training bases that have representative simulators for specialized undergraduate pilot training (SUPT), Euro-NATO Joint Jet Pilot Training (ENJJPT), and fighter aircraft flying training units (FTUs). The six bases visited allowed the research team to assess all the different training devices being used for the aircraft included in the study.

The objective of the simulator site visits was to evaluate the current state-of-the art for the simulators in AETC. The evaluation team constructed two evaluation tools to use in the evaluation. The first was a mission profile list called *AETC Simulator Evaluation Procedures*. This list contained the major mission areas the simulator operator would need to present during the evaluation mission such as takeoff, landing, night operations, and formation flying. This mission profile list was given to the technician or instructor operating the console. The second document was a simulator characteristic and technical performance checklist called the *Simulator Technology Assessment Questionnaire*. When filled out, it contained the physical description of the simulator system and the performance data observed, measured, or gathered by interview, during the pre-flight, mission flight, and post-mission debrief.

While a sizeable segment of the fleet of simulation devices in AETC is old by many standards, the architecture on which most of them are based (VME or PC) makes them potential candidates for upgrades that could be fairly easily integrated into existing training systems, thereby significantly reducing the cost to rebuild the devices. The technology throughout the fleet ranged from high fidelity cockpits with no visuals of any kind to trainers with full field-of-regard visual systems.

2.1 On-site Evaluation Procedures

Each simulator was evaluated using the same set of mission profiles and the performance evaluation checklist. The mission profile checklist given to the console operator is shown in Table 1. A brief explanation of the targeted evaluation is included.

Table 1. AETC Simulator Evaluation Procedures

AETC Simulator Evaluation Procedures	
Procedure	**Description**
1. Day VFR takeoff	During this phase, the ground environment was evaluated for scene fidelity, brightness, density, realism, and cultural features.
2. Pitch-out and landing and touch and go	Immediately following takeoff, the simulator was maneuvered for line-up on initial at pattern altitude and a pitchout to a touch-and-go was accomplished if the system had a visual. In those cases where the visual was of limited field-of-view, the downwind was flown to a radial and DME to begin the base to final turn in order to see if it was remotely possible, in a student environment, to use the simulator for transition training. This pass was also used to evaluate the richness of the database at the home field.
3. Climb to altitude and descend to low altitude for texture and scene evaluation	Following the touch and go, the simulator was zoomed to altitudes above 20,000 feet MSL. At altitude, the database was assessed for range ring settings, distance to the horizon, texture pattern interference, anti-aliasing, scene "popping", and level of fidelity, in general. Following this initial look, the aircraft was dived towards the ground to a very low level to check the performance of the image generator and the effects, if any, of resolution changes in the out-the-window scene during the descent.
4. Low altitude flying for system performance evaluation	At very low altitudes, the simulator was placed in maximum power and flown across the database to assess the performance of the visual system with regards to frame rate effects, loss of frame rate, ability to judge height, and the effects of high angular rate aileron rolls over the terrain, in general, and over the home station airfield with cultural features (3D) in the scene.
5. While low altitude test collision detection with terrain and features	Following the low altitude assessments in step 4 above, the simulator was descended into the ground to determine if collision detection was present. In most implementations when a collision with the ground occurs, the simulator should halt permanently or in temporary freeze state releasable by the IP. In higher fidelity simulations, these same characteristics should be present when colliding with cultural (3D) models attached to the terrain surface of the database. It should be noted that collision detection is also a feature that can be selected or de-selected, depending on the training objectives of the mission.
6. Rejoin with other target(s) if capable	If the simulator was capable of generating another target in flight, it was then inserted on the fly. If the system was not capable of doing this, the system was reinitialized with the second aircraft in the scenario. The target was then located visually or using available sensors and rejoined using the standard turning rejoin procedures or straight-ahead rejoins if the target could not easily be turned. The purpose of this phase was to evaluate the visual image for resolution, the effects of frame rate, the fidelity of the aircraft models, and the ability of the host and image generator systems to produce a stable, non-jittering target when close aboard.
7. Evaluate occulting visually	This task was intended to assess the whether or not another aircraft would

and in avionics, if applicable	disappear from sight visually when flying behind another cultural feature such as a building or behind a terrain feature such as a mountain or ridge line. In addition, if the simulator being evaluated had radar, the same test was to be accomplished in order to determine if the other aircraft would "disappear" electronically from the sensor being used to detect it. Due to the wide range of fidelity throughout the "fleet", the lack of the ability to easily maneuver computer generated entities in most of the AETC simulators, and the lack of networked targets to control in real-time, this area was assessed mainly by interviewing onsite or technical support center personnel.
8. Fly close formation with another aircraft, if able	Since most current systems have a reasonably high transport delay (time from stick input to measurable state change) and the flight aerodynamic model may not be exactly like the aircraft, this phase was not used to formally evaluate the ability of the simulator system to fly close formation. However, once aboard, the evaluator did fly in formation for a short period of time to assess other performance characteristics and to judge whether or not the system had the potential to operate in either a close, route, or extended trail formation environment.
9. Night approach	Upon completion of the day VFR tasks in Steps 1-8 above, the simulation was reconfigured on the fly or reinitialized to a night time environment. The first step was to evaluate the database fidelity, in general. That is, were there any visual clues to identify the ground environment, first, away from home field and, second, at the home field or primary night approach airfield? The evaluation also looked at the night sky to determine if any special features such as clouds and stars were present or available. Prior to leaving the night environment, a visual and an instrument approach were flown to the home field or practice instrument approach airfield. If the visual approach went well (no evaluator error), the instrument approach was accomplished with weather effects. The fidelity of the airfield environment in a night scene was also noted.
10. Experience special effects	If the simulation was capable, any special effects such as airburst, sun glint, lightning flashes, and missile smoke trails were implemented and evaluated. Most of the systems evaluated either had limited special effects available or none at all.
11. Air refuel, if capable	While no system in AETC had a fully operational air-to-air refueling simulation to include an articulated tanker boom, this portion of the evaluation consisted of rejoining on a tanker model, if available, or another aircraft such as the T-1, to observe the degree to which this phase of training could be simulated. Tanker (or other aircraft) fidelity and stability were two of the key parameters assessed.
12. Weather effects, if capable	During some point in the evaluation mission, a wide range of weather effects, as available, were instantiated and evaluated. This included general scene content when simply flying VFR, the ability to invoke cloud layers, the physical nature of the cloud layers (did they have thickness for example), the appearance of the weather effect, and the degree of simulation of fog and low visibility situations. To the best of our knowledge, no simulation system evaluated had the ability to present a physics-based environment where the weather phenomena simulated anything beyond just the visual effects. It should be noted at this point, however, that we know of NO real-time simulation used for training or research that does present a totally physics-based representation of the synthetic natural environment.
13. Full stop landing	After completion of the steps above, to the degree possible based on time available and system capability, the mission was terminated with a day VFR straight-in, full-stop landing. One last look was given to the airfield

	environment, the aerodynamic response of the host simulation system in the landing configuration, and the ability to land in the proper touch down zone and aero brake, if appropriate.
14. Discuss Instructor Operator Station	During mission execution, one of the AFRL evaluators was positioned at the instructor operator station (IOS) to observe and discuss the capabilities and ease of use of the IOS. Any further information needed by the flight evaluator was gathered by interview and demonstration following the simulator flight.
15. Discussions with contractors and users	Before leaving a site, an attempt was made to gather any technical data not easily available from discussion and gather any useful information pertinent to the technical characteristics or performance of those systems being evaluated. Where IPs (contractor or USAF) and/or students were available, the system just flown was discussed with them.

The *Simulator Technology Assessment Questionnaire* was used to gather the physical and performance characteristics of each AETC simulation system. The information collection was grouped into seven separate categories: System Identification, Visual Display System, Image Generator/Databases, Typical Problem Areas observed During Flight, Adequacy/Capability for Training Characteristics, Cockpit, and Instructor Operator Station. The physical characteristics were usually a fill-in-the-blank answer and included items like image-generator frame rates and display brightness. The performance data collected during an evaluation flight were answered with a simple "yes" or "no" assessment and looked for the existence of performance characteristics like double imaging or anti-aliasing.

The information collected on current training system architectures, features, capabilities, and performance characteristics helped the researchers determine an appropriate list of candidate technologies to be considered and evaluated by IPs later in the study.

3.0 Instructor Pilot Task Survey

The entire analysis of the study hinged on developing a comprehensive understanding of the current state of training in AETC and identification of the greatest training challenges or shortfalls. To gain this insight, the researchers constructed task surveys for each of the flying training courses in order to assess the current training programs and how well the existing training device technologies were meeting their needs.

Current instructor pilot (IP) involvement was essential for creating task surveys that were valid and understandable to each of the instructor populations. To help the researchers identify the appropriate flying tasks to use on the surveys as well as develop appropriate survey items, the researchers conducted a three-day workshop with current IPs from each of the different aircraft training courses being analyzed in this study. Table 2 summarizes the instructor qualifications and experience levels.

Table 2. IP Experience at Survey Development Workshop

Aircraft	Training Course	Base	Primary Aircraft	Military Flight Hours
T-6	SUPT	Moody AFB	B-52	4150
T-1	SUPT	Vance AFB	KC-135	2500
	SUPT	Laughlin AFB	C-5	2800
T-38A	SUPT	Columbus AFB	FAIP	600
	SUPT	Vance AFB	F-16	1100
	ENJJPT	Sheppard AFB	F-15	2500

AT-38B	IFF	Sheppard AFB	A-10	4216
T-38C	IFF (Pilot and WSO)	Moody AFB	F-15E	1200
F-15	B-Course	Tyndall	F-15	1100
	B-Course	Mesa Research Site (Retired F-15 pilot / IP)	F-15	3400
F-16	B-Course	Luke AFB	F-16	2300
	B-Course	Luke AFB	F-16	3400
	FAC-A, MANTIRN, NVG	Luke AFB	F-16	2600
			Average Hours	2451

After the survey development workshop, AFRL used its Internet-based survey tools to host the surveys on the Internet. The workshop IPs then had a chance to review and test the on-line version. To further ensure survey clarity, usability, and thoroughness the researchers traveled to various training bases to field-test the survey with small samples (at least two IPs) from each of the targeted survey populations. As a result of the field-testing, the wording of three of the questions and scales were clarified and the task lists for three aircraft were slightly modified.

Table 3 summarizes the final version of the five questions and rating scales that were used on the surveys for all of the different aircraft training courses included in this study.

Table 3. Task Survey Rating Scales

Question	Scale
1. How difficult is it for students to learn this task?	1 = Not difficult 2 3 = Difficult 4 5 = Extremely Difficult
2. How adequate is the amount of time currently allocated in the syllabus for the training of this task?	1 = Not enough 2 3 = Adequate 4 5 = Too much
3. How frequently does this task contribute to busted rides?	1 = Rarely 2 3 = Occasionally 4 5 = Frequently
4. How well trained are typical graduates of your course on this task?	1 = Barely proficient 2 3 = Proficient 4 5 = Mastered
5. How well do current training simulation devices prepare students for this task?	0 = No device available/applicable 1 = Poorly 2 3 = Adequately 4 5 = Extremely well

These survey questions were designed to collect information to help determine how well training is currently being conducted and identify those areas in which training is more difficult or problematic. The data collected from these surveys formed the basis for a priority weighting scheme. A priority weighting factor was created for each task that gave stronger weights to those tasks shown to be more difficult or problematic. Tasks with a higher priority weighting factor are greater targets of opportunity for enhancement by improved training simulation technologies. In the section on Quality Function Deployment, the use of the priority weighting factor will be further explained.

In addition to forming the basis for the technology investment prioritization, the survey collected a wealth of data that is expected to also be useful for those instructors and commanders making syllabus revisions or for helping decision makers understand the proficiency outputs of students in the flying training courses.

3.1 Data Collection

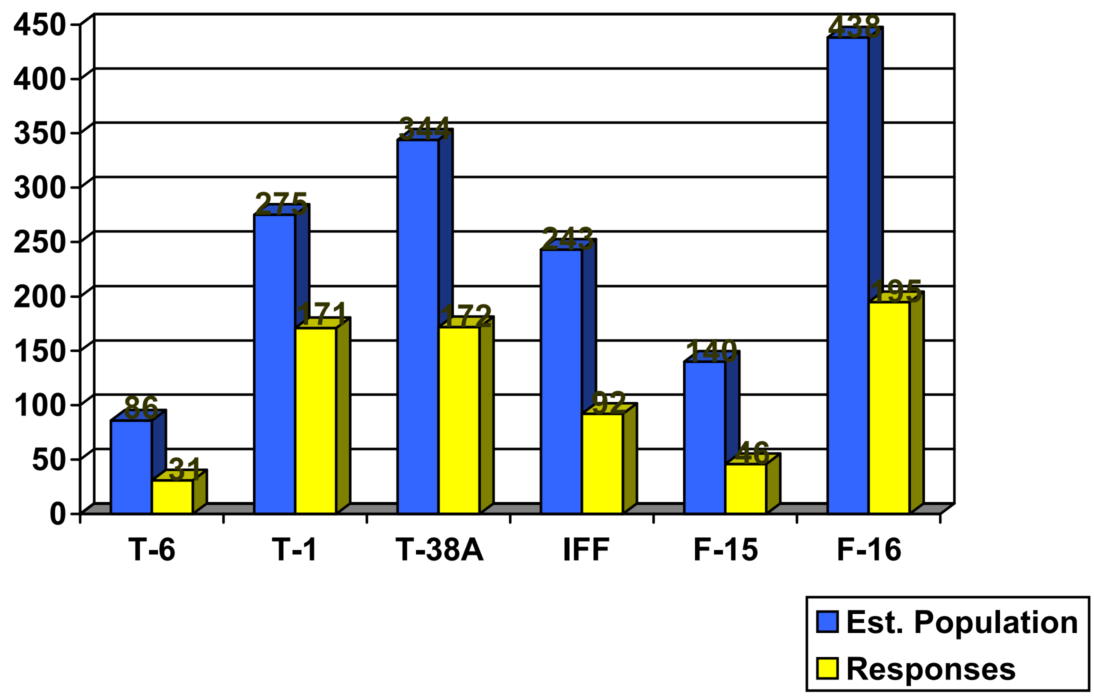
After field-testing of the surveys, they were administered on-line via the Internet to current IPs at each of the AETC SUPT, ENJJPT, IFF, and FTU bases. Table 4 lists the bases and training courses included in the study.

Table 4. Targeted Survey Populations

Base	Aircraft	Training Course
Laughlin AFB TX	T-38, T-1	SUPT
Vance AFB OK	T-38, T-1	JSUPT
Columbus AFB MS	T-38, T-1	SUPT
Sheppard AFB TX	T-38, AT-38B	ENJJPT, IFF
Moody AFB GA	T-6, T-38C	SUPT, IFF
Luke AFB AZ	F-16	B-course, NVG, FAC-A, and MANTIRN
Tucson ANG AZ	F-16	B-course, NVG
Springfield ANG OH	F-16	B-course, NVG
Kelly ANG TX	F-16	B-course, NVG
Tyndall AFB FL	F-15	B-course
Kingsley ANG OR	F-15	B-course, NVG

Official tasking came from the general officer level in AETC for all instructor pilots of the courses being studied to take the on-line survey. This kind of high-level interest and support was critical to achieve the kind of response rates we did and insure a valid data set. The Figure 1 shows the responses and estimated IP population sizes for each of the aircraft included in the study.

Figure 1. Survey Response Numbers



4.0 Quality Function Deployment

In order to assess the training technologies that could best be used in AETC, the researchers conducted a Quality Function Deployment (QFD) workshop with current IPs from each of the training aircraft being studied. When possible, the same IPs from the survey construction workshop were brought back for the QFD. Table 5 shows the experience of the IPs who participated in the QFD workshop.

Table 5. QFD Workshop Participants

Aircraft	Training Course	Base	Primary Aircraft	Military Flight Hours
T-6	SUPT	Moody AFB (USN pilot)	SH-60B	1500
	SUPT	Moody AFB	B-52	4150
T-1	SUPT	Vance AFB	KC-135	2500
	SUPT	Laughlin AFB	C-130	2000
	SUPT	Laughlin AFB	KC-135	2500
T-38A	SUPT	Columbus AFB	FAIP	600
	SUPT	Vance AFB	F-16	1100
T-38C	IFF (Pilot and WSO)	Moody AFB	F-15E	1200
F-15	B-Course	Tyndall	F-15	1800
	B-Course	Tyndall	F-15	1800
F-16	B-Course	Luke AFB	F-16	3400

	B-Course	Mesa Research Site (Reservist; former IP)	F-16	2800
	B-Course	Springfield ANG	F-16	2500
	FAC-A, MANTIRN, NVG	Luke AFB	F-16	2600
			Average Flight Hours:	2175

The purpose of the QFD was to have selected IPs score the usefulness of various simulation technologies to support the training of each task included in the survey of their aircraft. The 27 technologies were identified and defined by engineers and scientists from AFRL. They selected different simulation technologies that have the potential for being used for training various flying tasks. The findings from the simulator field evaluations helped the researchers define the candidate list of technologies. Some of the technologies already exist and are used in the field by only some of the simulators; others are newer technologies available today but are not being exploited yet; and a few are technologies that are not quite mature enough but should be in 2-3 years. To begin the QFD workshop, AFRL provided definitions, demonstrations, and discussions of the 27 different technologies to ensure that all the IPs had a common and thorough understanding of the technologies and their capabilities. The 27 different technologies scored in the QFD were as follows:

1. High Physical and Functional Fidelity Cockpit
2. Networked Simulators
3. Simulators Networked to Other Virtual Assets
4. Motion Platform
5. Force Cueing Devices
6. Cockpit Environmental Sounds
7. PC-Based Part-Task Trainer
8. PC-Based Part-Task Trainer, with Voice Interaction
9. PC-Based Part-Task Trainer, with Intelligent Tutoring
10. Visual Systems with 20/20 Visual Acuity
11. 72° Horizontal/62° Vertical Field-of-View Visual Display
12. 220° (Horiz) FOV Visual Display
13. 360° (Horiz) FOV (or as limited by aircraft design) Visual Display
14. Head-Mounted Display (a.k.a. Virtual Reality)
15. Eye-Tracking Technology
16. Photo-Realistic Visual Database
17. Accurate Sensor Simulations
18. Realistic Weather and Atmospheric Effects
19. 3-D Ground Models
20. Computer Generated Moving Models
21. Computer Generated Interactive Entities
22. Computer Generated Adversaries
23. High Fidelity Threat and Electronic Combat Environment
24. Mission Planning System Integration
25. Digital Debriefing System
26. Virtual Reality Brief/Debrief Tool
27. Current Part-Task Trainers

To conduct the QFD, the IPs broke into small groups based on their aircraft types with AFRL facilitators in each group. IFF, F-15, and F-16 IPs were in the same workshop group because the majority of their training tasks are the same across platforms. The IPs were instructed to score the tasks individually and then discuss their ratings within the group to then arrive at a consensus score. Forcing the groups to arrive at a consensus score, rather than just calculating an average score, drove a great deal of discussion and exploration amongst the participants. This discussion brought out different ideas on technology application that might not have been considered or adequately factored into the scores had the groups been allowed to use a more expedient averaging process.

IPs were instructed to score each technology based on the following scale of usefulness:

1 – Not at all useful
2 –
3 – Somewhat useful
4 –
5 – Very useful

They were also instructed to score the technology independent of cost and to ignore whether they already had the technology at their base or not. The raw QFD scores should just reflect how useful a technology would be for training a given task. A cost-benefit analysis is not a part of this analysis model. It is up to the modernization planners and system acquisition personnel to work with the contractors to determine estimated costs and assess the benefits. The results of this analysis show the planners how that technology can best be used for the greatest gains.

4.1 QFD Results and Technology Prioritization

The technology usefulness scores arrived at by consensus for each task during the QFD were then multiplied by a priority weighting factor. The priority weighting factor was derived from the data collected with the task survey described above that was given to all IPs across AETC. The priority score was created by:

- finding the mean IP rating for each task on each question (e.g., all of the IPs rated Task X on Question 1; the mean rating of Task X on Question 1 was calculated)**

- adding the mean ratings of each task from questions 1, 2, 4, and 5. Question 3, which asked IPs to rate how often a task contributed to busted rides, was not included in the calculation of the priority index because the AFRL team thought that the question was too speculative in nature and their was little coherence in the data collected.

For example, if 4 IPs were asked to answer questions 1, 2, 3, and 4 (Q1, Q2, Q3, and Q4) about tasks 1, 2, and 3, the resulting raw data would look like this:

	TASK 1				TASK 2				TASK 3			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
IP1	2	2	3	2	2	1	1	3	2	3	2	3
IP2	2	3	4	3	3	2	2	3	4	4	3	4
IP3	3	4	1	4	4	3	3	3	3	2	4	5
IP4	3	3	1	3	3	2	2	1	3	1	1	4

Next, the mean rating for each question on each task is calculated:

	TASK 1				TASK 2				TASK 3			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
AVG	2.5	3	2.5	3	3	2	2	2.5	3	2.5	2.5	4

**NOTE: Before the means were derived, the ratings collected in the survey process were rescaled so that each one was directionally congruent; on some of the rating scales, a "5" was a positive attribute and on other questions a "5" was a negative attribute. The ratings were rescaled so that higher ratings indicated more problematic tasks.

Then, the mean ratings for each question are added together to give a priority weighting factor:

	TASK 1	TASK 2	TASK 3
PRIORITY WEIGHTING	2.5+3+2.5+3= 11	3+2+2+2.5= 9.5	3+2.5+2.5+4= 12

The higher the priority weight, the more problematic a task is and the greater the target of opportunity for training enhancement with advanced simulation technology. In our example above, Task 3 would be the greatest target of opportunity for training enhancement with advanced simulation technology.

Each of the raw technology usefulness scores assigned by IPs during the QFD were then multiplied by the priority weighting factor for each task, thereby creating weighted priority scores that are composed of a technology usefulness score (from the QFD) and task ratings (from the IP task survey). These weighted priority scores for each technology were then added to yield a total weighted priority score for each technology. A higher score indicates that a technology has a greater potential for positive training impact for that course.

Continuing our example, if 3 training technologies (TECHNOLOGY 1, TECHNOLOGY 2, and TECHNOLOGY 3) were assigned usefulness scores during a QFD, the technology usefulness scores for each task would be multiplied by the priority weighting factor for each task; the resulting weighted priority scores for each technology would then be summed to give a total weighted priority score for each technology:

		TECHNOLOGY 1		TECHNOLOGY 2		TECHNOLOGY 3	
	Priority Weight Factor	Technology Usefulness Score (from QFD workshop)	Weighted Priority Score (priority score X technology usefulness score)	Technology Usefulness Score	Weighted Priority Score	Technology Usefulness Score	Weighted Priority Score
TASK 1	11	2	11*2 = 22	4	44	1	11
TASK 2	9.5	3	9.5*3 = 28.5	4	38	3	28.5
TASK 3	12	4	12*4 = 48	5	60	4	48
TOTAL WEIGHTED PRIORITY SCORE			**22+28.5+ 48= 98.5**		**142**		**87.5**

In the example, Technology 2 looks to have the greatest potential positive impact on training.

The technologies were then ordered from highest to lowest total scores. This rank order can help to identify the technologies that may have the greatest impact for improving pilot training. A cautionary note: the resultant technology priority lists are only ordinal in nature. That is, technologies with higher total weighted priority scores have a greater potential impact on training. However, a technology with a total weighted priority score of 400 does NOT have twice as much potential impact on training as a technology with a total weighted priority score of 200. In addition, total weighted priority scores should not be compared across platforms. Task surveys for different planes contained different numbers of tasks; and since the total weighted priority scores are products of the number of tasks on a survey, planes with surveys containing greater numbers of tasks would likely have overall higher total weighted priority scores for technologies.

5.0 Conclusion

No other time in history have we seen as much readily available technology for training warfighters as we see today. The analytic process described in this paper provides a useful way of identifying those training technologies with the greatest opportunity for enhancing flying training programs. Critical to this entire process was involvement by the instructor pilots themselves. Not only did the survey responses from over 700 instructors form the basis for the priority weighting, the instructor participation in the survey development and technology scoring workshops was critical. This process has yielded a great amount of useful data and recommendations for the leadership in the USAF Air Education and Training Command. The involvement by current instructors in the workshop as well as the analytic basis from instructor surveys yields great validity and credibility to the analysis and recommendations. The information from this process will form the basis for their simulation roadmap and guide their technology investments to enhance the training of U.S. Air Force pilots.

Aircrew Mission Training via Distributed Simulation Progress in NATO

Barry Tomlinson
Training & Simulation Business Group
Future Systems Technology Division
QinetiQ Ltd
The Enclave
Thurleigh
Bedford MK44 2FQ
United Kingdom
+44 (0) 1234 22 5377 (Tel.)
+44 (0) 1234 22 5409 (Fax)
bntomlinson@mail.qinetiq.com

Chairman, NATO MTDS Task Group
(SAS-034/MSG-001)

Jan van Geest
TNO Physics and Electronics Laboratory
P.O. Box 96864
2509 JG The Hague
The Netherlands
+31 70 374 0210 (Tel.)
+31 70 374 0652 (Fax)
vanGeest@fel.tno.nl

Chairman, NATO MSG-001 Task Group

Key Words:
Mission Training, Distributed Simulation, Networked Training, Collective Training, Coalition Training

Summary

Mission Training via Distributed Simulation (MTDS) exploits modern simulation technology to provide a new concept in aircrew collective training for air operations. MTDS has been the subject of NATO RTO activities since 1998. The MTDS Task Group is working to create a prototype NATO synthetic MTDS environment, conduct a multi-national COMAO exercise and assess its potential to support training to enhance NATO's operational effectiveness in multi-national air operations. This activity is known as "Exercise First WAVE" (Warfighter Alliance in a Virtual Environment), the first ever multi-national wide area networked real-time simulation of Combined Air Operations conducted in NATO. This paper reviews the work of the NATO MTDS Task Group, examines the nature and potential of Aircrew Mission Training via Distributed Simulation and reports progress with "Exercise First WAVE". It summarises lessons learned so far about defining and creating a distributed training environment and about the technical and training issues which need to be addressed in order to implement and exploit MTDS in NATO and the nations.

1 Introduction

1.1 MTDS – the Background

MTDS – Mission Training via Distributed Simulation – is the NATO name for a new concept in aircrew mission training. MTDS exploits modern simulation technology to improve operational readiness and contribute to mission success. Over the past five years MTDS has been the subject of several NATO RTO activities (Tomlinson, 2000, 2002, Van Geest & Tomlinson, 2002, 2003). A Military Applications Study (SAS-013) was undertaken from 1998-2000 to "assess the potential of advanced distributed simulation to complement live flying training in order to enhance NATO capability to conduct combined air operations". The MTDS Task Group was then initiated in May 2001 to build on this preliminary work. A NATO RTO Symposium (SAS-038), held in Brussels, Belgium, in April 2002, was devoted to "Air Mission Training Through Distributed Simulation (MTDS) – Achieving and Maintaining Readiness".

Document Identifier QINETIQ/FST/CP034263
hfm_mtds_paper_16.doc

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

MTDS focuses on "collective training" and inter-team skills and applies to aircrew and mission crew - pilots, navigators, and all weapons and mission system operators. They are assumed to possess the basic individual and team skills needed to be categorised as "combat ready". Such aircrew must then master the collective skills necessary in multinational air operations as part of a larger unit involving two or more teams from two or more countries. The definition of collective training applicable to the air domain, composed by the initial SAS-013 study, is

> ***"Collective training involves two or more 'teams', where each team fulfils different 'roles', training to interoperate in an environment defined by a common set of training objectives."***

Collective training is thus concerned with multi-role, multi-platform interactions, co-ordination and communication.

MTDS is about creating a distributed training environment which immerses aircrews and mission crews in a realistic operational scenario and enables them to conduct collective mission training for air operations. MTDS exploits modern advanced computing and networking technologies to bring together a federation of crewed aircraft and mission simulators in a synthetic environment. Crews perform the end-to-end COMAO (Composite Air Operation) processes of planning, briefing, mission execution and debriefing in the distributed training environment surrounding the federation. MTDS is feasible technically and has the potential to offer expanded training opportunities for development of collective skills.

1.2 The NATO MTDS Task Group

The NATO MTDS Task Group combines the SAS-034 Task Group sponsored by the Studies, Analysis and Simulation (SAS) Panel and the MSG-001 Task Group sponsored by the NATO Modelling & Simulation Group. This Task Group is currently working to create a prototype NATO synthetic MTDS environment, conduct a Multi-National COMAO exercise and assess its potential to support training to enhance NATO's operational effectiveness in multi-national air operations. This simulation-based aircrew training exercise, known as "Exercise First WAVE" (Warfighter Alliance in a Virtual Environment), will be the first ever multi-national wide area networked real-time simulation of Combined Air Operations conducted in NATO. The exercise, planned for September 2004, explores issues of matching training requirements and technical capability, and exposes the need for a multi-national exercise development team to address these aspects.

The overall objectives of the current MTDS Task Group are:

- To demonstrate, investigate and assess the potential of MTDS to enhance aircrew mission readiness for NATO coalition operations, including training, simulation technology and management aspects;
- To increase awareness of MTDS capabilities amongst national and NATO military staffs;
- To establish a set of guidelines, procedures and standards based on experience and lessons learned;
- To propose further actions needed to implement and exploit MTDS in NATO and the nations.

1.3 Organisation

Members of the task group are Air Forces, Industry and Research and Technology organisations from Canada, France, Germany, Italy, the Netherlands, the United Kingdom and the United States of America.

Within the MTDS task group, Exercise First WAVE is managed by a Steering Group, supported at the working level by five specialist task teams:

- Operations and Training
- Technical
- Security
- Assessment
- Awareness

Experienced military officers lead the Operations and Training team, to ensure that objectives appropriate to the operator community are defined and met. The steering group is made up of the National representatives of the participating countries, and the leaders of the task teams. A full-time programme manager, sponsored by the USA, helps the Task Group chairman manage the entire effort. More information about the technical aspects of First WAVE is given in a companion paper (Cerutti & Greschke, 2003).

2 Training Need

2.1 Air Operations

In modern air operations, a mission package typically employs platforms of many kinds, not only in the essential strike and escort roles, but also in such other crucial support roles as electronic warfare (EW), suppression of enemy air defence (SEAD), command & control (AWACS) and air-air refuelling (AAR). They all contribute collectively to mission success, and their crews need to train together.

A further feature of modern air operations is that the balance between flying skills and weapon system operation is evolving to place greater emphasis on sensor manipulation, information management, situation awareness, decision making and communication. Such a change in emphasis generates new training requirements for a complex tactical context in which sensor and weapon suites can be fully employed in association with other aircraft. Future operations will employ a “system of systems” in which single aircraft will themselves be part of an operational network, in what is referred to as “network-centric warfare” or “network-enabled capability”. To ensure that aircrew can employ their systems effectively, and be well co-ordinated with others, they need opportunities to train with all appropriate and relevant assets.

2.2 Live Flying Training and the Training Gap

NATO training for combined air operations is accomplished today solely through live flying training, such as the Tactical Leadership Programme (TLP), and through exercises, such as the NATO Air Meet (NAM).

Opportunities for NATO aircrews to participate in TLP and complex air exercises are limited. NAM takes place only once a year, and then only for a relatively small number of crews. An individual crew or pilot may participate only rarely in Red Flag. As mission complexity increases, and as the trend towards network-centric, and network-enabled, operations gathers momentum, there is an increasing need for combat assets to practise operational integration with ISTAR and other support assets. A training gap is developing and new training methods are needed to provide increasingly complex scenarios.

This increasing complexity means that the full range of mission training requirements cannot be fulfilled in the air. Among the many factors limiting the capability to conduct realistic mission training during live flying training exercises are:

- Increasing pressure to reduce training costs
- Restrictions on airspace and lack of adequate training ranges
- Operational factors such as improved weapons system performance capabilities, and security constraints on the use of electronic warfare systems
- Mission complexity and rules of engagement (ROE)
- Environmental and safety restrictions, including the inability to fire weapons or use chaff/flare
- An unrepresentative mission environment, with no threats that fire back

Modern weapon system performance capabilities and the growth of data links are extending the "tactical reach" of an air package. Thus, aircrew combat training in the 2000-2010 time frame will need to change from the training of the 1990s, with emphasis now on higher order weapons system employment skills requiring co-ordination, communication, and complex judgement.

Developments in simulation technologies, and in the ability to simulate an immersive and convincing mission environment, offer ways to fill this emerging training gap. Some initiatives equivalent to NAM and Red Flag, but based on simulation and MTDS, will need to be established.

3 Technical Capability

3.1 Simulation technologies

Simulation technologies have advanced substantially over the past 5 years: computing power continues to improve, communication networks have become widely available (and also more affordable) and satellite imagery has enabled enhanced terrain databases for (almost) anywhere in the world to be created. Computer generated forces also are available (and, indeed, are necessary) to augment piloted platforms to provide additional Blue force assets, as well as the reactive Red air and ground-based defence necessary for tactical realism. These developments have made it possible to build a realistic and stressful synthetic battlespace in which numerous participants, no matter where they are located, can operate together in a shared mission. Such technology has been evolving steadily, and has now reached the stage where it is being used regularly in some nations for squadron training (e.g. the US DMT programme). Multi-national experiments have also been conducted in a research context.

Modern advanced simulation is very different from the traditional kind. The new generation of simulators can offer mission-related capability that enables them to participate in MTDS. To be able to join an MTDS exercise, simulators need to be both "mission-capable" and "network-capable". However, it must be realised that not all today's "legacy" simulators are suitable for use in MTDS.

Mission-capable means manned simulators must have sufficient fidelity to allow pilots to "train as they intend to fight" and perform their mission tasks in a realistic manner. This requires that the aircraft cockpit, flight performance, sensors and weapons be simulated to an appropriate (moderately high) level of physical and functional fidelity, and that the mission environment provides a sense of "immersion".

Network-capable means each manned simulator must be capable of exchanging data in a secure manner with other participating simulations, using an agreed, standard protocol. The participating simulators must also be interoperable with one another within a common synthetic environment such that the cause and effect relationships (interactions) in the synthetic world correspond to the cause and effect relationships in the real world.

3.2 Training environment

A training environment to exploit MTDS needs more than just capable simulation assets, in the form of operational multi-ship training facilities provided by the nations. It needs an extensive "MTDS training system", with facilities to support the training audience (the crews) in all phases of the complete operational cycle, from issue of the Air Tasking Order, via planning and briefing, to flying the mission (in simulators) and finishing with post-mission debriefing. In addition, the simulators must be capable of meeting the exercise management and data capture requirements identified by the training manager.

Key elements of the distributed training environment are illustrated in Figure 1. All these elements are directed at supporting the training objectives for a particular exercise.

Tactical Environment

The Tactical Environment defines the characteristics of, and performance for, all tactical entities which interact with each other and with the Synthetic Natural Environment. The tactical environment consists of manned (virtual) simulators and computer-generated forces. With some exceptions, the blue forces are manned simulators, and the red forces are CGF.

Synthetic Natural Environment

The Synthetic Natural Environment (SNE) represents the geophysical environment of the mission space, i.e. terrain and natural features, as well as 3D cultural features, together with the atmosphere and the weather. The SNE has to be represented sufficiently well, in the visual system and in how it influences sensors and weapons, to serve the training objectives of the exercise, and to provide a maximum sense of immersion to the pilots.

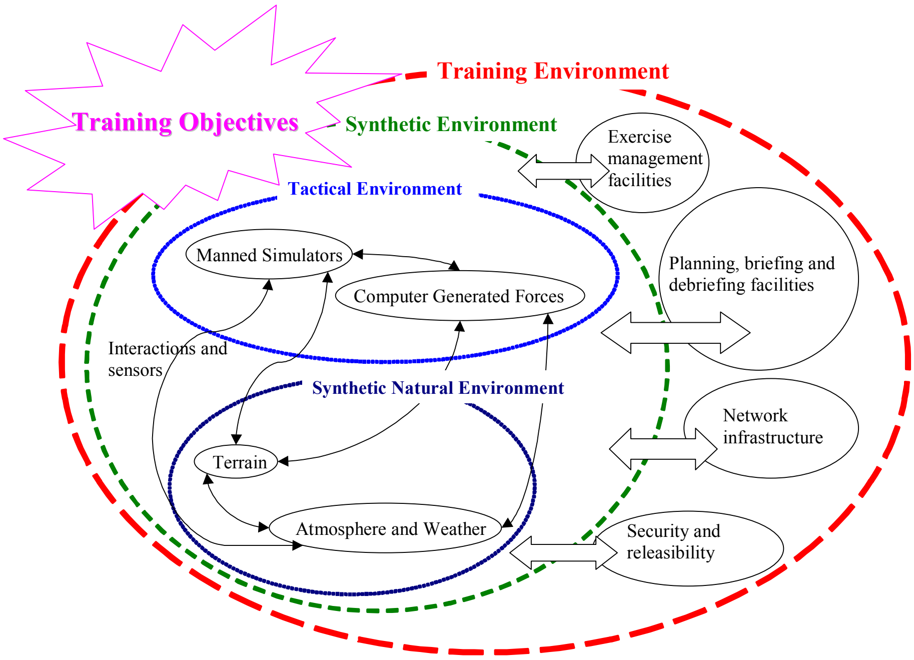


Figure 1 Elements of the multinational mission training environment

Interactions and Sensors

Interactions and sensors are a fundamental element of the distributed training environment. They provide the mechanism that enables one entity to know about the existence and behaviour of another entity. Interactions exist between the different tactical entities, and between a tactical entity and the geophysical environment. Achieving "fair play" and realistic interactions are crucial to the success of a networked training exercise. Simulators will differ in the level of modelling and simulation of their sensors and weapons. The capabilities of each networked simulator will determine its role and degree of participation in an exercise. False results must be avoided.

Exercise management facilities

Setting-up, controlling and using a network of simulations as a training tool are significant activities. An exercise scenario that facilitates the training objectives has to be designed, implemented and tested. The personnel that assure that the training process proceeds according to the requirements are often referred to as the White Force.

One way the white force can influence the training process is by the injection of trigger events into the scenario. Such events are intended to provoke blue force interactions and promote training in the area of mission critical or time critical targeting. An example of a trigger event is the appearance of a Surface-to-Air missile system from a hidden location in the scenario to feed a real-time targeting loop. The trigger events will be planned in advance.

A discussion of some experiences obtained from a research programme into the application of networked simulators to collective training (Smith & McIntyre, 2003) suggests there are major challenges in conducting aircrew collective training using simulators and crews at geographically dispersed sites.

Mission planning, briefing and debriefing

During the training period, the environment must allow the trainees to do all planning, briefing, execution and debriefing activities necessary to receive maximum benefit from the COMAO mission. The individual sites should therefore be connected not only during the execution, but also during the planning, briefing and debriefing activities. For planning, the sites will use their organic planning facilities. Briefing may be facilitated using interconnected interactive whiteboards and voice, so that the mission commander can brief all crews at the same time.

Network

An obvious requirement for a distributed training environment is the need for a wide area network with sufficient bandwidth and a low latency to support real-time man-in-the-loop simulation and all the data that needs to be shared. This item also includes portals or other devices that may have to be used to connect systems together that use a different interoperability standard, e.g. DIS or HLA.

Security and releasability

Most MTDS exercises will be classified. This implies that the participating sites should be accredited for the applicable classification level, and that all data exchange between the sites will be encrypted. Furthermore, a project agreement between the participating nations has to be put in place in order to be able to release classified data over the network.

Security procedures require considerable time to establish. Therefore, a sound security plan is one of the first products that has to be produced for any exercise including networked classified simulations.

4 Exercise First WAVE

4.1 Objectives

In support of the overall objectives of the MTDS Task Group listed earlier, the top-level aim of Exercise First WAVE is to create a distributed simulation environment (the Federation[1]) in which warfighters can conduct a Composite Air Operation in order to assess the potential of NATO MTDS. Exercise First WAVE will bring together national training systems and enable them to operate in the same simulated scenario.

As is usual for any exercise, a set of top level training objectives has been defined:

- To practise daytime COMAO procedures employing fighter escort/sweep, AAR, SEAD, reconnaissance and AEW in a hostile EW environment
- To exercise procedures for defensive operations with Fighter Areas of Responsibility (FAORs) and point defence tasking
- To employ EW resources in support of offensive and defensive air operations
- To plan and integrate a multi-national COMAO in a defined threat environment
- To brief a COMAO package generated from dispersed locations
- To conduct mission debriefs
- To engender efficient team-working skills between Nations and differing elements of the COMAO package
- To develop a tactical appreciation of real-world threats
- To expose aircrew to situations to which they would not normally encounter in a peacetime environment
- To establish lessons identified

[1] Federation is part of the language of the High Level Architecture, used as a basis for development of Exercise First WAVE.

These establish what needs to be included in the synthetic environment and the nature of the scenario. It is assumed that all participating aircrew and mission crews are already trained to "combat ready" status and therefore have the skills required to work together as a small team. It is the integration of individual teams with other elements of the package that is the training focus of the exercise. The work in the US on Mission Essential Competencies (Colegrove & Alliger, 2002) will be applied and extended to coalition air operations.

4.2 Scenario

Exercise First WAVE will represent a scenario which offers a geographically realistic location and a complex operational environment. A substantial Blue Force, consisting of about 30 manned simulators, together with additional assets represented by computer driven forces, will undertake strike missions, supported by AWACS and AAR. These missions are modelled on Composite Air Operations (COMAO), as illustrated conceptually in the diagram below (*Figure 2*). Most of the participating simulators will be operational training systems (such as an F-15 4-ship from the Mission Training Center at Eglin AFB in the USA), flown by operational, combat ready, crews, while a few will be R&D simulators (but also flown by operational pilots). A Red Force will consist of both manned and computer driven air defence assets, together with an integrated ground-based air defence system. The entire networked synthetic environment will operate at a classified level, open only to participating nations.

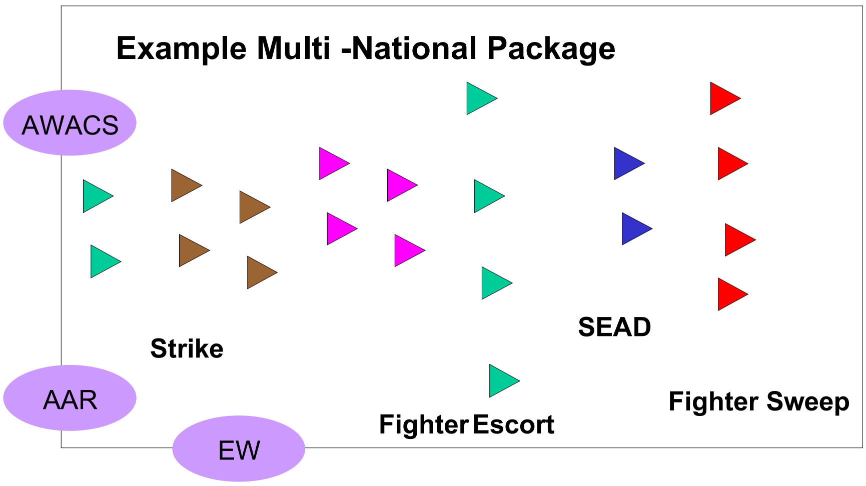


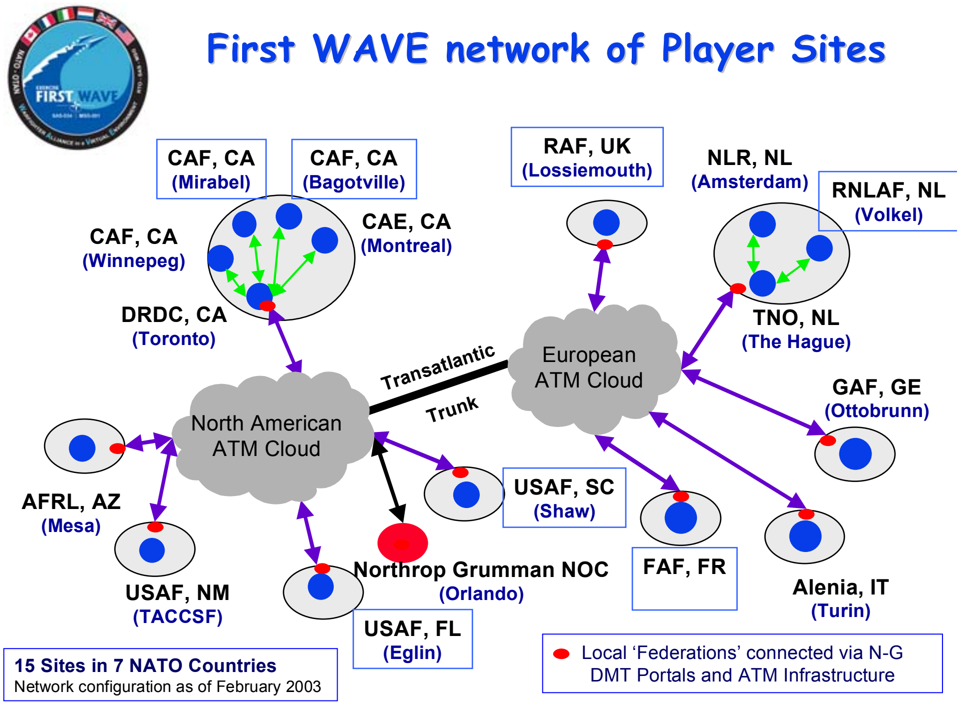
Figure 2 Example package structure

4.3 Conduct

Given that the simulated air operations will be focused on aircrew training potential, with "suitable" levels of command and control (AWACS, CAOC), they will be conducted in an operationally realistic manner. Following an initial familiarisation day, each of the three mission days will start with the release of the Air Tasking Order (ATO). Under the leadership of a Package Commander selected from the participants, each formation will then plan and brief their tasks, in co-ordination with other formations as necessary, fly their sorties and then conduct subsequent formation and mass debriefs. Ensuring crew immersion by following such a complete mission cycle is one of the key goals of the exercise, as defined in the User Requirements Document. It is also vital that all participants receive a training experience of value to *them*. From this outline, it is clear that the exercise will be much more than just a federation of simulators. In fact, it will be a Distributed Training Environment. How well crews are able to plan and brief/debrief with others at dispersed sites will be one of the key topics to be assessed. Comprehensive and robust communications facilities are important, for both non-flying and flying phases. During the flying phase, realistic simulation of multiple radio channels is a critical element in creating a believable representation of the tactical environment.

4.4 Network

Exercise First WAVE will involve simulators and training facilities at about 15 sites in 7 NATO countries (*Figure 3*). Setting up and managing the wide area network (WAN) is itself a major activity. We are fortunate to have the support of Northrop Grumman Mission Systems to do this. They are the US contractors



already responsible for the operations and integration aspects of the US DMT programme.

Figure 3 First WAVE Network

4.5 Assessment & Evaluation

The assessment team is responsible for the assessment of the potential operational training benefits of MTDS for COMAO training and for documenting the strengths and weaknesses identified through the execution of Exercise First WAVE. The mission essential competencies mentioned earlier will be used as a foundation for identifying training gaps and for assessing the training potential of MTDS.

5 Challenges in Training & Technology, & Lessons Learned

5.1 Training

The first challenge is to define and agree a set of training objectives. These need to ensure that ***all*** participants will receive a training experience of value to them. However, the agreed training objectives also have to match the training assets available, and recognise specific national objectives as well as coalition goals. In Exercise First WAVE we had a group of participating nations, who offered assets and we had to build a scenario to exercise them all. It would help in future to have a single NATO or coalition "General" with a requirement, who says something like "I want to bring together a coalition package from nations X, Y, Z, … in order to practise and improve capabilities P, Q, R, … in scenario F." What you want to do and who you want to do it with have major influences on the overall feasibility, from both a technical and security point of view.

How missions will be conducted and managed needs to be defined and agreed, in order to specify and scope the key features of the total distributed training environment. The exercise management team also needs to be identified. These are the people who will supervise and manage the exercise to ensure the training objectives are met. A scenario and associated trigger events need to be defined. This requires careful pre-planning and preparation. Planning, briefing and debriefing facilities and requirements also need to be identified. A large part of the training value in *co-located* live exercises has been found to stem from face-to-face planning, briefing and debriefing. The same value is needed in MTDS, with the extra challenges of being geographically distributed.

5.2 Technology

Integration of dissimilar (legacy) simulators into a distributed training environment raises several interoperability issues, including:

- Simulators need to be mission-capable: the degree to which each simulator represents the capabilities of the real aircraft, its sensors and weapons will influence the role it can play in the mission;
- Simulators need to be network-capable: each simulator must be able to exchange and interpret data in a secure manner with other participating simulators, using an agreed standard protocol;
- Simulators have to work together in a compatible manner and support the types of interaction between sensors, weapons and defensive aids required by the scenario;
- Security and releasability: the training environment has to provide a good balance between maximum training value and national data release constraints;
- Synthetic Natural Environment: in particular correlated visual and sensor databases and weather representation are not easy to achieve over dissimilar legacy simulators; however, to meet collective training objectives, these issues may have lesser importance.

While a convincing and worthwhile simulated mission environment can be created, there are still technology limitations that will constrain the range of missions that can be performed and the range of collective skills that can be fully exercised. There is a need for continuing, and co-ordinated, research in many areas:

- to improve our understanding of collective mission training needs in a multi-national context, the nature of team fluidity and the competencies that need to be trained
- to understand the balance between live and synthetic training, and how best to exploit the strengths of each type of training
- to establish effective methods and tools for the conduct of mission training exercises from a number of geographically separate locations and to help understand the relative benefits of "distributed" and "co-located" simulator facilities
- to develop methods to relate training objectives to necessary simulator functionality, to design MTDS events, to assess performance, and to evaluate multi-national training programmes for collective skill training
- to improve the capabilities of computer generated forces
- to identify the key factors in achieving interoperability among mission simulators of different design, including network issues, modelling of interactions and external control and to investigate whether a NATO Air FOM needs to be developed
- improved representation of the environment, to include effects of bad weather and night and their influence on sensors ands weapons
- representation of data links

5.3 Lessons Learned

Exercise First WAVE is still being planned and developed, so many of the lessons learned so far are about defining and creating a distributed training environment and about organisational issues. It is essential to keep at the forefront that the principal objective is about training, not technology, even though solving technical issues is essential. A key lesson is that MTDS is about more than just simulators.

The first lesson learned is that many challenges exist of an organisational nature, as well as in the training and technology areas. The scope of Exercise First WAVE is larger than any collaborative geographically distributed simulation programme attempted before. A "Memorandum of Understanding" (MOU) or similar arrangement is needed. Even there, issues arise, as one country may find MOU an acceptable label, but another seeks a "Technical Arrangement".

The potential "show-stopper" is security. Security and information disclosure issues need to be recognised as having a major impact on the future of MTDS. A "Security Processes & Procedures" document is essential as the basis for accreditation. Any computer-based system handling classified data has to achieve accreditation by appropriate national authorities. To be part of a multi-national network, it also has to be approved by a multi-national authority. Who this should be was not obvious when we began, but after some false starts we have found that the appropriate body is the Multi-National Security Accreditation Board (MSAB). When satisfied, the MSAB will issue a certificate enabling connection to take place. However, this does not deal with information disclosure, which is essentially a *national* decision about the data each nation is willing to release.

Among other challenges and lessons is that of finance. The Task Group has no central budget but relies on national funding. Furthermore, the parties involved include industry as well as government. Many of the actual training systems are owned by industry and used by Air Forces on a "pay by the hour" basis. Significant contributions are being made to Exercise First WAVE by industry using their own money.

5.4 Success criteria

The MTDS Steering Group has defined a set of goals and success criteria under three headings:

Training

- that Exercise First WAVE provides crews with a credible and worthwhile training experience
- that Exercise First WAVE demonstrates the ability to conduct realistic coalition aircrew training from brief to execution to debrief
- that data and experiences can be gathered to properly assess the potential of MTDS in NATO

Security

- that processes and agreements for multi-national secure real time MTDS networks can be defined and agreed
- that a secure synthetic coalition training environment can be established and delivered

Legacy/Infrastructure

- that Exercise First WAVE provides a proof of concept for MTDS and distributed mission operations that will be the cornerstone of training transformation
- that a documented infrastructure and lessons are left for future use
- that relations and links among national staffs are established for future C2 and fighter mission rehearsal/training

Achievements to date have included the preparation of the User Requirements Document, the Management Plan, and the generation of a Test Federate. The work of the Task Group has been of significant value already to Nations, in terms of both policy & technical understanding. Dissemination of the concept of MTDS is actively pursued by briefings to NATO and national organisations.

6 Future potential of Aircrew Mission Training via Distributed Simulation

6.1 MTDS in NATO

MTDS is feasible now, but all stakeholders have much to learn and need to acquire more experience in defining, setting up and running MTDS-based training exercises. Currently, it is the NATO R&D community that is exploring MTDS in terms of both training benefits and technical challenges. "Exercise First WAVE" will, hopefully, be a major milestone in taking MTDS forward. To go further, we need to "hand it on" to a clearly identified "owner" or "champion" of MTDS in the NATO operational community.

This champion would take responsibility for how best to explore and exploit what MTDS can offer, and would establish a NATO team to initiate MTDS-style exercises. Such a team needs to include and integrate expertise from the NATO Military Commands, from the NATO AEW&C Force Command (unique in NATO in actually owning relevant simulator assets), from the Tactical Leadership Programme and from national Air Warfare Centres.

We need to advance our understanding, through practical experience, of how all participants in such training exercises (exercise management team/white force, aircrew and the technical team) can effectively work together to achieve defined training objectives. Significant effort is required to define training objectives, to plan and prepare the scenario and to conduct the exercise. Experience to date suggests that every such exercise will furnish lessons learned and provide insight into fresh areas to explore. Strong liaison between the operational and training research communities will need to be maintained, to ensure that operational lessons learned and training needs are understood and that implementation of MTDS makes best use of scientific knowledge.

An evolutionary programme of MTDS exercises should be adopted, building up progressively over a few years from relatively small-scale training events to larger exercises. Issues to be considered include:

- training requirements and objectives (what needs to be trained, with whom and how often)
- representative scenarios
- suitable configurations (federations) of simulation facilities
- methods and tools to help manage distributed training exercises
- a NATO network infrastructure which is manageable and flexible
- security processes and procedures

Our experience so far with MTDS is in fast jet operations. The implications of using MTDS for training in other kinds of air operation, and in joint operations, also need to be examined.

7 Concluding Remarks

MTDS is an important application of advanced technology to military training and can potentially provide a major contribution to coalition readiness and effectiveness in future NATO air operations. The training opportunities that MTDS can offer – a complex tactical environment and few (if any) restrictions – will be increasingly vital in an era of Network-Centric Warfare and Network-Enabled Capability. The federation being developed for First WAVE, and the guidelines that emerge from the exercise, will be a basis for further implementation and exploitation of MTDS within NATO.

Many nations are acquiring training simulators which are potential participants in future MTDS exercises, and have research programmes to define national use of networked simulation for training. In some countries operational aircrew already use this type of training. National (and NATO) procurement processes need to ensure that simulator-based training systems are acquired with the appropriate mission-capable and network-capable features.

NATO has been exploring MTDS and its potential since 1998 through an initial study (SAS-013), a major symposium (SAS-038) and now through the MTDS Task Group. It would help communication and understanding if all groups in NATO use MTDS as the appropriate label in a NATO context.

The MTDS Task Group will deliver, in September 2004, Exercise First WAVE, the first ever multi-national wide area networked real-time simulation of Combined Air Operations conducted in NATO. This will bring together more than 30 manned simulators in seven nations in a realistic and representative scenario. Apart from Exercise First WAVE itself, deliverables will include lessons learned, guidelines on implementation and a proposed plan for future exploitation. Although the exercise itself will have controlled access, for security reasons, the main results will be reported to all NATO nations.

Enabling Exercise First WAVE to happen is consuming a lot of effort, organisationally to achieve appropriate formal agreements and agree security processes; and technically to solve such challenges as the provision of a secure network, the supply of encryption devices and the integration of a variety of national training systems into a unified whole. For MTDS to be usable in a more "every day" manner in future the organisational processes must be simplified and new technical challenges resolved speedily.

Security is the potential show-stopper. To bring together any viable coalition-based collection of aircrew training systems will require a secure wide area network. Achieving accreditation for such a network involves national and international processes. Exercise First WAVE has broken new ground here and has hopefully established how such a network can be authorised in future, via the Multi-national Security Accreditation Board. Nations still have to resolve their own data disclosure policy.

To strengthen the capability and enable the full potential of MTDS to be realised in NATO (and the nations) needs continued national and collaborative research in such areas as:

- creating a distributed mission training environment which serves the whole plan-brief-fly-debrief cycle
- briefing and debriefing methods and tools
- representation of data links
- interoperability
- computer generated forces
- improved representation of the environment, to include effects of bad weather and night and their influence on sensors ands weapons

We need to sustain the momentum behind MTDS. Exercise First WAVE will take place in September 2004. Enabling this Exercise to happen has brought together the Air Forces and other agencies of seven nations, together with substantial contributions from industry. This momentum should not be lost. We would hope that the military community in NATO, at both Allied Command Operations (based at SHAPE) and the new Allied Command Transformation (formerly SACLANT), will take the concept forward and enable "Second WAVE" (or whatever it might be called) to take place soon. There are other nations who wish to participate. A steady, and evolving, sequence of exercises is needed. We all still have a lot to learn about how to create and conduct such simulator-based exercises to achieve real training benefits.

We are making good progress towards achieving Exercise First WAVE and hope it will contribute to the future success of MTDS in NATO.

References

Cerutti S & Greschke D (2003). *'Aircrew Mission Training via Distributed Simulation (MTDS) - Development of the Multi-Country Complex Synthetic Environment'*, paper 18 NATO Human Factors and Medicine Panel Symposium on "Advanced Technologies for Military Training", 13-15 October 2003, Genoa, Italy.

Colegrove C M & Alliger G M (2002). *'Mission Essential Competencies: Defining Combat Mission Readiness in a Novel Way'*, Paper presented at NATO RTO Symposium on "Air Mission Training Through Distributed Simulation (MTDS) – Achieving and Maintaining Readiness", Brussels, Belgium, 3-5 April 2002

Smith E & McIntyre H (2003). *'Distributed Mission Training – How Distributed Should It Be?'*, paper 17 NATO Human Factors and Medicine Panel Symposium on "Advanced Technologies for Military Training", 13-15 October 2003, Genoa, Italy.

Tomlinson B N (2000). *'Aircrew Mission Training via Distributed Simulation – a NATO Study',* presented at NATO Modelling & Simulation Conference, RMCS Shrivenham, UK, 24-26 October 2000. Published in NATO RTO MP-071

Tomlinson B N (2002). *'Aircrew Mission Training via Distributed Simulation – a Review of NATO initiatives',* Paper presented at NATO RTO Symposium on "Air Mission Training Through Distributed Simulation (MTDS) – Achieving and Maintaining Readiness", Brussels, Belgium, 3-5 April 2002

Van Geest J & Tomlinson B N (2002). '*Mission Training through Distributed Simulation -Contributing to Warfighter Integration*', presented at NATO Modelling & Simulation Conference, Paris, France, 24-25 October 2002.

Van Geest J & Tomlinson B N (2003). *'A Multinational Mission Training Capability for NATO Composite Air Operations'*, International Conference, Exhibition and Network for Training, Education and Simulation (ITEC), London, 29 April – 1 May 2003

Author Biography

Barry Tomlinson was Study Director (1998-2000) for the Military Applications Study SAS-013 on "Aircrew Mission Training via Distributed Simulation" sponsored by the NATO Studies, Analysis and Simulation Panel and is at present chairman of the MTDS Task Group (SAS-034/MSG-001) on "Mission Training via Distributed Simulation (MTDS) - Concept Development and Demonstration". Currently employed at QinetiQ (formerly the Defence Evaluation & Research Agency) at Bedford in the UK, he has more than 35 years' experience in the technology and application of flight simulation and synthetic environments to aircraft and systems development and to training.

Jan van Geest has contributed to NATO MTDS activities since 1998 and currently represents both MSG-001 and The Netherlands in the First WAVE Steering Group. He is employed at the Physics and Electronics Laboratory of TNO (Netherlands Organisation for Applied Research, TNO-FEL) in The Hague, NL and advises the Royal Netherlands Air Force on the use of simulation for mission training. His experience includes organisational and technical definition of debriefing for live and simulated collective exercises.

List of acronyms

AAR	Air-Air Refuelling
AEW	Airborne Early Warning
ATO	Air Tasking Order
AWACS	Airborne Warning and Control System
C2	Command and Control
CAOC	Combined Air Operations Centre
DMT	Distributed Mission Training
EW	Electronic Warfare
FAOR	Fighter Area of Responsibility
FOM	Federation Object Model
ISTAR	Intelligence, Surveillance, Target Acquisition and Reconnaissance
MSAB	Multi-National Security Accreditation Board
MTDS	Mission Training via Distributed Simulation
NAEWFC	NATO AEW&C Force Command
NAM	NATO Air Meet
NATO	North Atlantic Treaty Organisation
NMSG	NATO Modelling & Simulation Group
RTO	Research & Technology Organisation
SACLANT	Supreme Allied Commander Atlantic
SAS	Studies, Analysis and Simulation (Panel of the RTO)
SEAD	Suppression of Enemy Air Defence
SHAPE	Supreme Headquarters Allied Powers Europe
TLP	Tactical Leadership Programme

Distributed Mission Training – How Distributed Should It Be?

Ms Ebb Smith, BSc
Defence Science and Technology Laboratory
Rm 126, Bldg 115
Bedford Technology Park
Thurleigh, Bedfordshire
MK41 6AE
United Kingdom
E-mail: mesmith@dstl.gov.uk

Ms Heather McIntyre, MSc
Defence Science and Technology Laboratory
Rm 126, Bldg 115
Bedford Technology Park
Thurleigh, Bedfordshire
MK41 6AE
United Kingdom
Email: hmmcintyre@dstl.gov.uk

Summary

The UK's MoD has funded a programme of applied research[1] to explore the benefits to be gained from using networks of simulator, or Synthetic Training Environments (STEs), for multi-role air mission training (i.e collective training[2]). Within the UK the use of networked simulation in this context has become known as Mission Training through Distributed Simulation (MTDS). Within the US it is known as Distributed Mission training (DMT). The Defence and Science Technology Laboratory (Dstl) and QinetiQ have undertaken the MoD sponsored research via a series of trials. The trials have been conducted under the banner heading of RAPTORS[3]. To date four trials have taken place; Ebb and Flow, SyCOE, VirtEgo and SyCLONE. All have been conducted using the synthetic Composite Air Operation (COMAO) test-bed created specifically to assess the potential of MTDS. Combat-ready, front-line aircrew and an expert White Force from the UK's Air Warfare Centre (AWC) Tactical Wing and Training have participated in all four trials. The research has indicated that there is, potentially, much to be gained from the use of networked simulation for MTDS. The question remains as to the extent that participants should or could be distributed during MTDS exercises. This is particularly pertinent if the aspiration is to use networked simulation for coalition training, because, of necessity this would require some training participants to be geographically dispersed. The last two trials therefore included a Wide Area Network (WAN) to link together research facilities in Canada, the UK and US. This paper will discuss the outcome of these trials with particular reference to SyCLONE.

Overview of VirtEgo and SyCLONE

Trial VirtEgo[4] took place in November 2001, whilst Trial SyCLONE[5] was undertaken in January 2003. The UK elements of the trials were funded by MoD and Strike Command, with the collaborative aspects carried out under the auspices of The Technical Co-operation Panel[6] (TTCP). As in the previous trials, VirtEgo and

[1] Under the sponsorship of the Director of Equipment Capability (Theatre Airspace)

[2] 'Collective mission training' is defined as two or more teams training to interoperate in an environment defined by a common set of collective mission training objectives, where each team fulfils a different military role. NATO SAS-013 Study

[3] Research into Aircrew Performance and Teamwork using Operationally Realistic Scenarios

4 VirtEgo - Virt stands for virtual and Ego a conscious thinking subject

5 SyCLONE - Synthetic CoaLition Operation in a Networked Environment

6 A long-term UK/US/CA collaborative project is being developed to take this research forward under the auspices of two groups of the Technical Co-operation Panel (TTCP); the Human Resources and Performance Group, Technical Panel 2 (HUM TP-2, Training Technology) and the Aerospace Systems Group, Technical Panel 1 (AER TP-1 Aerospace Operational Analysis and Simulation).

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held in Genoa, Italy, 13 – 15 October 2003, and published in RTO-MP-HFM-101.

SyCLONE were designed to execute emulation of a COMAO and incorporated all mission phases: tasking, brainstorming/planning, briefing, mission execution and debriefing [Ref 1]. The trials were modelled on live collective training exercise such as the NATO Tactical Leadership Programme (TLP) and UK Tactical Leadership Training (TLT) and the US Red Flag exercises. The major caveat being that the live collective training exercises have participants co-located. For elements of VirtEgo and SyCLONE, participants were geographically dispersed. This was a critical aspect of the research.

The synthetic environment created for the VirtEgo and SyCLONE enabled operational aircrew, based in the UK, Canada, and US to perform a COMAO based training exercise. For all missions, the rules of engagement (ROE), and special instructions (SPINS) were based on those of the theatre in question and the missions were designed to enable the aircrew to fly in accordance with coalition tactics and doctrine.

For all four of the RAPTORS trials, a team from the AWC Tactical Wing provided an expert White Force (WF). They provided a critical element of the exercise management function. They also provided expert assessment of collective performance throughout each mission day. This mirrors the role they perform in facilitating the UK's national air collective training exercises.

The manned participants in the trials formed part of a small package of aircraft flying a coalition mission within a shared synthetic battlespace. Other elements of the package, specifically Suppression of Enemy Air Defence (SEAD) assets, were represented during mission execution by computer generated forces (CGF). During planning the SEAD element lead was represented by a WF role player. For each mission day one of the participating aircrew was selected to be the Package Commander (PC) responsible for co-ordinating the package's efforts in order to achieve commander's intent.

The package formed part of a much bigger offensive air support operation, with other coalition missions being flown against hostile ground and air threats within the same timeframe. The added complexity of greater numbers of friendly and hostile forces was provided largely by CGF, with the addition of some manned air threats.

VirtEgo

Trial VirtEgo was the first trial to combine both training and research thrusts. VirtEgo was designed as a 'proof of concept' for preparing front-line Qualified Weapons Instructor (QWI) students for COMAOs in readiness for the two week operational phase (Ops Phase) of their CQWI training programme. An encrypted trans-atlantic WAN to Air Force Research Laboratory (AFRL) in Mesa, Arizona, US was also implemented to allow a coalition package to fly together in a shared synthetic battlespace. The WAN included a 'stealth' link between AFRL and Defence Research and Development (DRDC) Laboratories, Toronto, Canada. An initial distributed planning, briefing and debriefing was also provided. This capability utilised commercial-off-the shelf (COTS) technology and took the form of a video-teleconferencing (VTC) system and interactive whiteboards i.e. a SmartBoard™.

The SmartBoard™ provides a touch sensitive screen, of relatively large physical area, which is convenient for multiple users. The Smartboard™ technology allows users to write on the board as though it were a dry marker board; these inputs can then be transmitted via Microsoft NetMeeting to other networked SmartBoard™ . Electronic files created at one location can also be shared, thus for example, a PowerPoint presentation can be viewed simultaneously at different locations.

The findings from VirtEgo suggested that it was possible to combine training and research objectives within one trial, but that the technology must first be robust and proven. [Ref 2]. A perceived utility in pre-deployment training for ab-initio crews and general COMAO refresher training was also found. However, the ability of the STE to support mission training for dispersed participants was not proven. There were technical difficulties and front-line crews did not fly the AFRL simulators. This may have biased the findings.

SyCLONE

From the UK perspective, Trial SyCLONE was designed to explore the impact of distributed versus colocated mission training on the UK participants (aircrew and WF). In previous trials the UK players had been colocated[7] [Ref 3]. The coalition aspects of the research were maintained via long-haul transatlantic links to both AFRL in the US and Defence Research and Defence Canada, Toronto in Canada. The manned elements comprised:

For Trial SyCLONE the manned participants were:

- 1 x UK Ground Attack (GA) 4-ship (Jaguars)
- 1 x UK Air-to Air (A-A) 4-ship (Tornado F3s)
- 1 x US Swing role 4-ship (F-16Cs)
- 1 x Canadian GA 4-ship[8] (CF-18s).

The training research focus in SyCLONE was on the effects of distribution on planning and co-ordination and ultimately on execution of the mission itself. While simulation is deemed to be essential to explore the effectiveness of the mission plan, it was not the focus of training research in this trial. Aircrew may already have some experience of distributed briefing/debriefing in military exercises, such as Cope Thunder. Collaborative planning of complex missions is not something that aircrew normally carried out in a distributed way during operational training.

The importance of the planning phase can often be overlooked in simulation exercises when the emphasis tends to be on the mission and the time spent in the simulator. However, anecdotal evidence suggests that up to 80 to 90% of the benefit of participation in large scale, complex exercises, whether live or simulated, comes from involvement in the planning process. The stated objective of planning is to generate a product, namely the plan. Really one of the most important benefits of planning is the provision of a rigorous, structured way to learn about a problem space and therefore to develop judgement [Ref 4].

During trial SyCLONE, UK crews planned, briefed, debriefed their missions in separate locations but at the same site. As in VirtEgo they were using technology specifically provided for the purpose, namely VTC and SmartBoards™ to communicate with each other. The WF was also kept separate from the aircrews. This meant that in effect the distributed STE comprised five dispersed nodes: three at Bedford UK, one at AFRL US and one at DRDC Canada. A schematic is given in Figure 1.

[7] Little previous research has been done on the impact of distribution on teamwork and team effectiveness in real world task environments. A comprehensive review of teamwork research, estimated that of more than 4,000 team studies, 95% were to do with colocated teams performing tasks in laboratory conditions. From the remainder, very little can be deduced about the consequences of distribution for effective real world training.

[8] The CF-18 4-ship comprised two man-in-the loop players and 2 CGF wing-men.

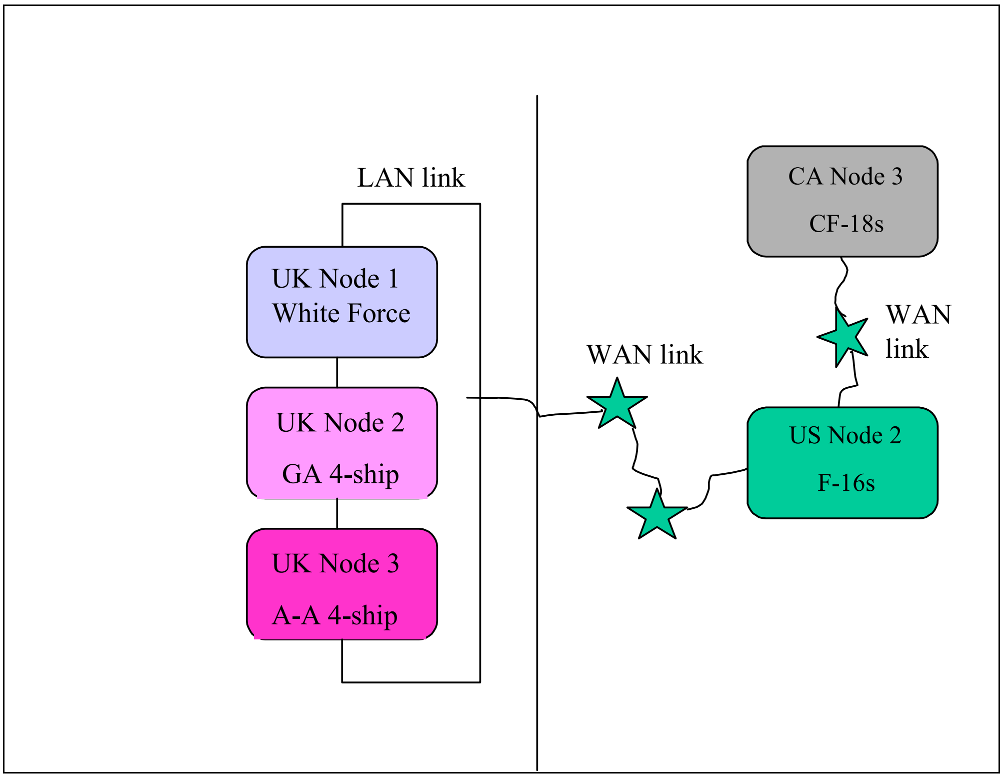


Figure 1: Experimental set-up with UK forces distributed

SyCLONE – findings

Analysis was conducted using a variety of assessment methods and tools. These included questionnaires, interviews and extensive use of subject matter expert opinion (the WF). A preliminary analysis of the data indicted that SyCLONE did not provide as fully optimised a training environment for the UK participants as did the previous three trials. There were a number of variables and confounding factors that contributed to this outcome. These included:

- Late mission days for UK crews
- Late tasking of aircrew to participate in the trial
- Current ops tempo placed an additional burden on WF
- Use of new technology, particularly for distributed planning, briefing and debriefing
- Some of the more specific findings and analysis are now discussed in turn.

Distributed planning, briefing and debriefing: The research team developed a very comprehensive concept of operation (conops) for distributed Planning, Briefing and Debriefing activities (PBD). The results of questionnaires given to aircrew indicated that they were happy with the instructions given on how to use the technology but did not feel the PBD system was user-friendly or robust.

The results also indicated that dispersion had impacted upon training value with the greatest effect being felt during the planning phase, resulting in a more 'simplistic' plan being produced. This is not necessarily a desirable outcome and certainly does not exploit some of the more obvious benefits of a complex MTDS environment. On the distributed days, problems were also experienced by the WF who felt unable to function as effectively when dispersed from crews. They believed that their ability to monitor and where appropriate guide the planning process was compromised.

The overall opinion by UK participants was that whilst the PBD technology was usable it detracted from the training benefits. The preference as in previous trials was for colocated synthetic COMAO training, irrespective of whether or not it is representative of real world operations.

To summarise:

- Distributed planning resulted in a simple and less integrated plan which did not fully exploit the capabilities of the other package elements
- The WF felt the Package Commander (PC) was not able to communicate effectively with other members of package
- The WF felt the technology too cumbersome for planning process
- Aircrew observed not as fully immersed in 'total experience' as in previous trials. No real 'buy in' to the synthetic experience, difficult to maintain suspension of disbelief through all phases, hence more critical of the simulation than in previous trials
- Low WF Situational Awareness (SA) on aircrew planning process, resulted in an inability to fully optimise the training environment.

As discussed, the UK crews did manage to have one day colocated where they were able to plan together. The WF provided an assessment of the aircrews performance during planning, briefing and debriefing for each of the four mission days. A summary of this is given below in Table 1 and clearly indicates that the day on which the crews were colocated (mission day four) produced a more comprehensive, well thought out and tactically considered plan.

Table 1: White Force assessment of Trial SyCLONE Planning Phases over Four Mission Days

Mission 1 (distributed)	Mission 2 (distributed)	Mission 3 (distributed but colocated for debrief)	Mission 4 (colocated)
	Planning		
• Not able to assess planning phase adequately • PC not able to communicate effectively with other package members • Ist day was perceived as battle with unfamiliar technology	• Not able to assess planning process adequately • Inadequate technology to support overview of distributed planning • PC couldn't communicate effectively with other package members • Technology seemed to impact on ability of aircrew to conduct workup & produce sophisticated plans. Much frustration with the technology • Result is a simpler plan – Force Flow type approach - not able to develop an integrated detailed plan	• Not able to assess planning phase adequately • Same problems as previous days – very little SA on aircrew planning process but slight improvement – more bandwidth & better audio	• Able to assess planning for UK crews. UK planning resulted in better plan than on previous days – more detailed in plan, more tactical thought, using work rounds to overcome tech difficulties to distributed US & CA nodes. • Use of VTC & SmartBoards™ affected distributed planning process – technology too cumbersome to use in normal planning process

It is well understood that a good plan makes for a better mission. The WF view was borne out by subsequent, in-depth post-trial analysis of a number of key events that occurred during SyCLONE's final mission day[9]. Aircrew comments also indicated that this day was preferred to the previous three mission days.

These observations support the findings from previous trials. Previous experience of colocated trials shows that rapport begins to develop between colocated crews after two days together; this is accompanied by a measurable improvement in performance. The effects are illustrated in Figure 2; this figure shows some of the results from an earlier RAPTORS trial (trial SyCOE[10]) which took place in January 2001 [Ref 5]. The data shown are derived from an Assessment Criteria questionnaire, which asks for assessments on thirty-one criteria of effective mission performance. These criteria are shown in table 2.

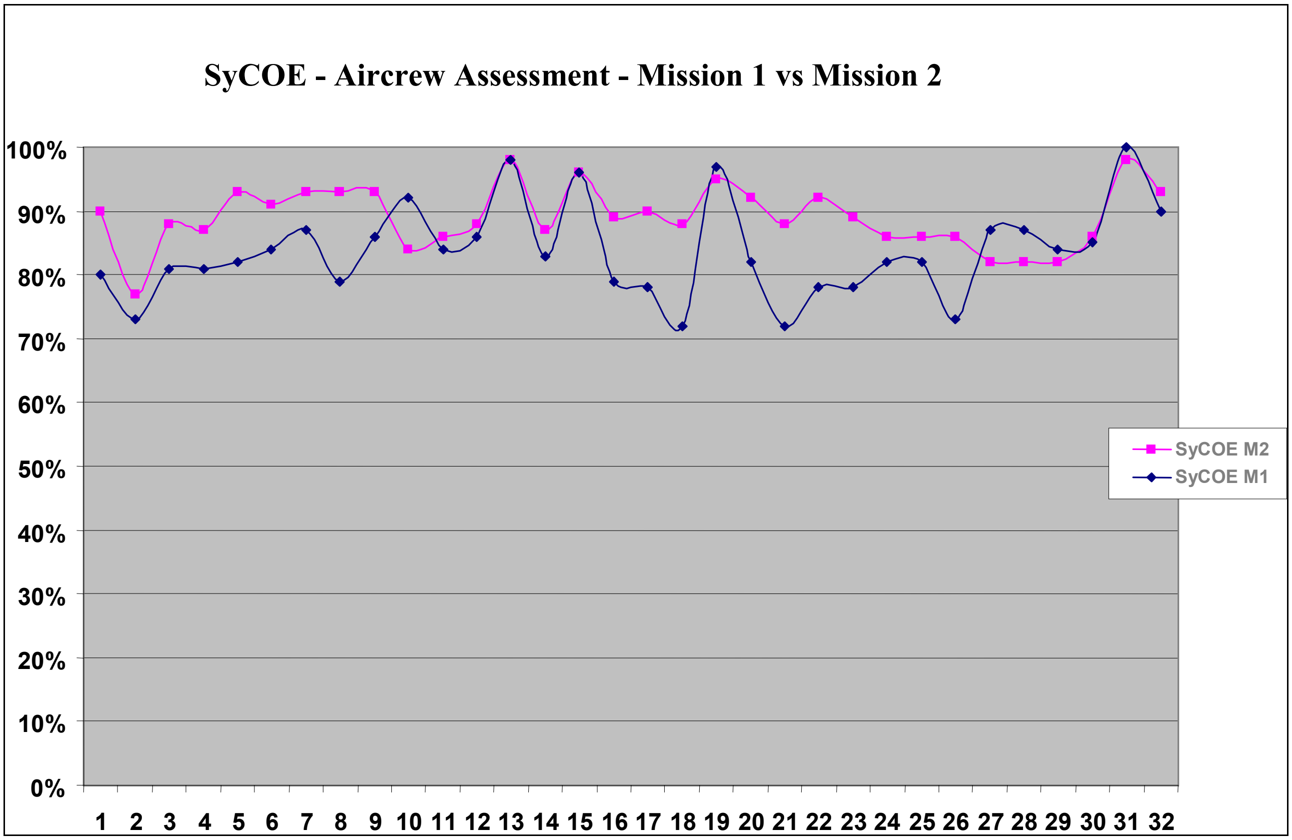


Figure 2: SyCOE aircrew assessment – mission 1 vs mission 2

Aircrew performance started from a good baseline in mission 1 and improved significantly in mission 2. This improvement was particularly marked for three areas:

- appropriate review of tactics as a result of lessons learnt in previous mission,

[9] This analysis is being undertaken by QinetiQ as part of an MoD funded study entitled *Quantifying the Effectiveness of Collective Training*.

[10] The training design of trial SyCOE provided a more appropriate comparison to SyCLONE than VirtEGO

- how well aircrew understood and implemented the briefed operational procedures
- 'between formation' awareness of other team's capabilities

Figure 3 shows results from trial SyCLONE.

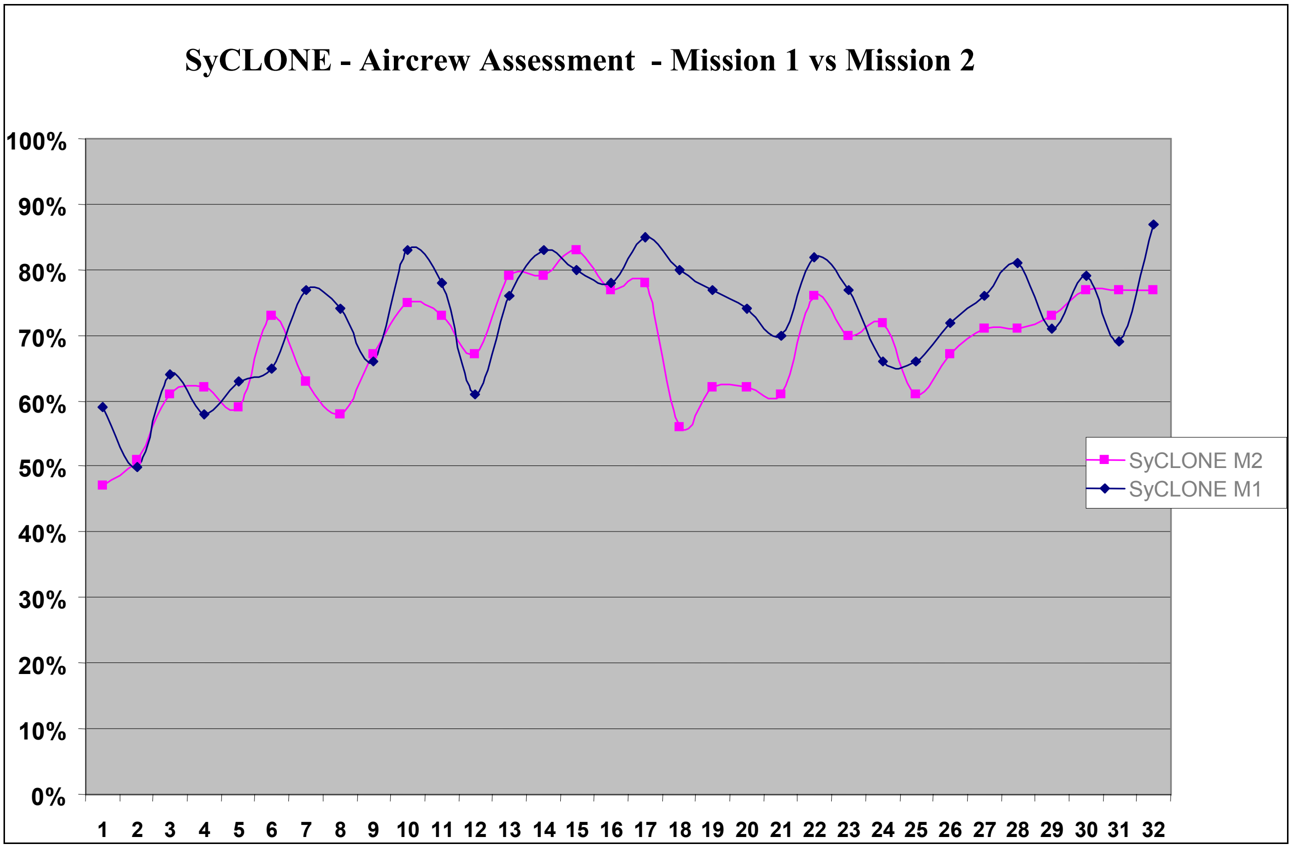


Figure 3: SyCLONE aircrew assessment – mission 1 vs mission 2

Aircrew performance during mission 1 was not given a particularly high rating, as compared to mission 1 in SyCOE. No obvious improvements in aircrew performance can be observed in SyCLONE mission 2. In fact there appears to be a decrement in performance in some areas, particularly:

- appropriate review of tactics as a result of lessons learnt,
- effectiveness of the plan,
- how well each role appeared to understand where they fitted into the 'bigger picture' of the COMAO

Table 2: Mission effectiveness - performance assessment criteria

Performance Assessment Criteria	
01	How effective was the plan?
02	How comprehensive was the brief ?
03	To what extent were all eventualities explored and addressed during the planning phase?
04	How effective were the tactics employed during the mission?
05	Were relevant lessons learnt and actions thoroughly debriefed?
06	How well was the expertise available during the planning phase, utilised by the PC?
07	How well were the needs, workload and time constraints of the other roles taken into account by the PC during planning?
08	How well did each role appear to understand where they fitted into the 'bigger picture' of the COMAO?
09	How effective were the responses to injects and self-generated problems during the mission?
10	How aware were the aircrew of relevant events and problems that could impinge on the mission
11	Given the knowledge available to them at the time how well did aircrew response to events & problems that could impinge upon the mission?
12	How well was formation integrity maintained throughout the mission?
13	How good was the comms discipline within the formations?
14	How effective were the formation leads in co-ordinating the assets within their formations?
15	How good was the comms discipline between formations?
16	How effective was the PC in co-ordinating the assets within the package?
17	How well were correct tactics employed against SAM or A-A threats?
18	How appropriate were any review of tactics made as a result of lessons learnt?
19	How well were the role-specific tasks demonstrated?
20	How well did the aircrew understand and implement the briefed operational procedures?
21	To what extent was the overall effectiveness of the group positively influenced by the PC leadership?
22	To what extent did the PC appear to benefit from the experience of command?
23	How appropriately did the others respond to the PC leadership style?
24	To what extent were the appropriate responses made to in-flight injects and self-generated problems?
25	To what extent was appropriate flexibility demonstrated?
26	How much confidence to individuals appear to have in their own capabilities?
27	Within formations, how much awareness did element leaders and their no. 2s appear to have of each others capabilities?
28	To what extent did formation members appear to be confident enough to rely on each others actions?
29	Between formations, how much awareness did teams appear to have of each others capabilities?
30	To what extent did the different formations appear to be confident enough to rely on each others actions?
31	To what extent were the overall objectives of the mission achieved?

Trust and confidence: Collective training exercises, whether live or synthetic, should endeavour to support the development of inter-role trust and confidence. This is an essential component of interoperability in theatre. To quote:

'The secret of a successful air campaign is interoperability
and the most important component of interoperability is trust'
Air Cdre Stu Peach, Commandant Air Warfare Centre

An apparent inability of UK and US crews to develop trust via good inter-team cohesion and build rapport was observed during SyCLONE. A questionnaire given to UK aircrews asked them to provide trust and confidence ratings for the other roles that were participating in the trial. It also asked them to rate the level of confidence they had in the ability of these other players to contribute to mission success. Pre and post trial ratings of trust and confidence, made by UK aircrew, were analysed and are presented in Figures 4 and 5. Figure 4 shows pre and post trial ratings for trust while Figure 5 shows pre and post trial ratings for confidence.

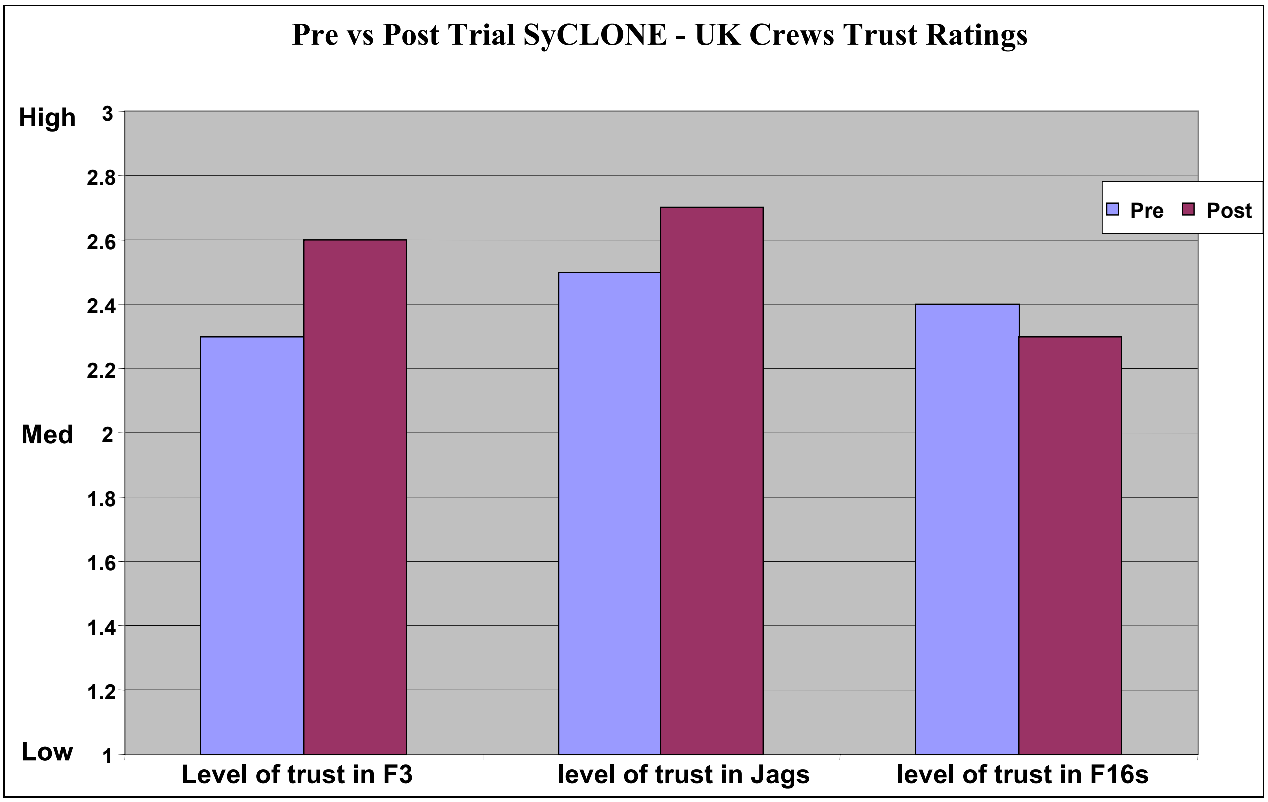


Figure 4: Pre v Post trial UK aircrew ratings for trust

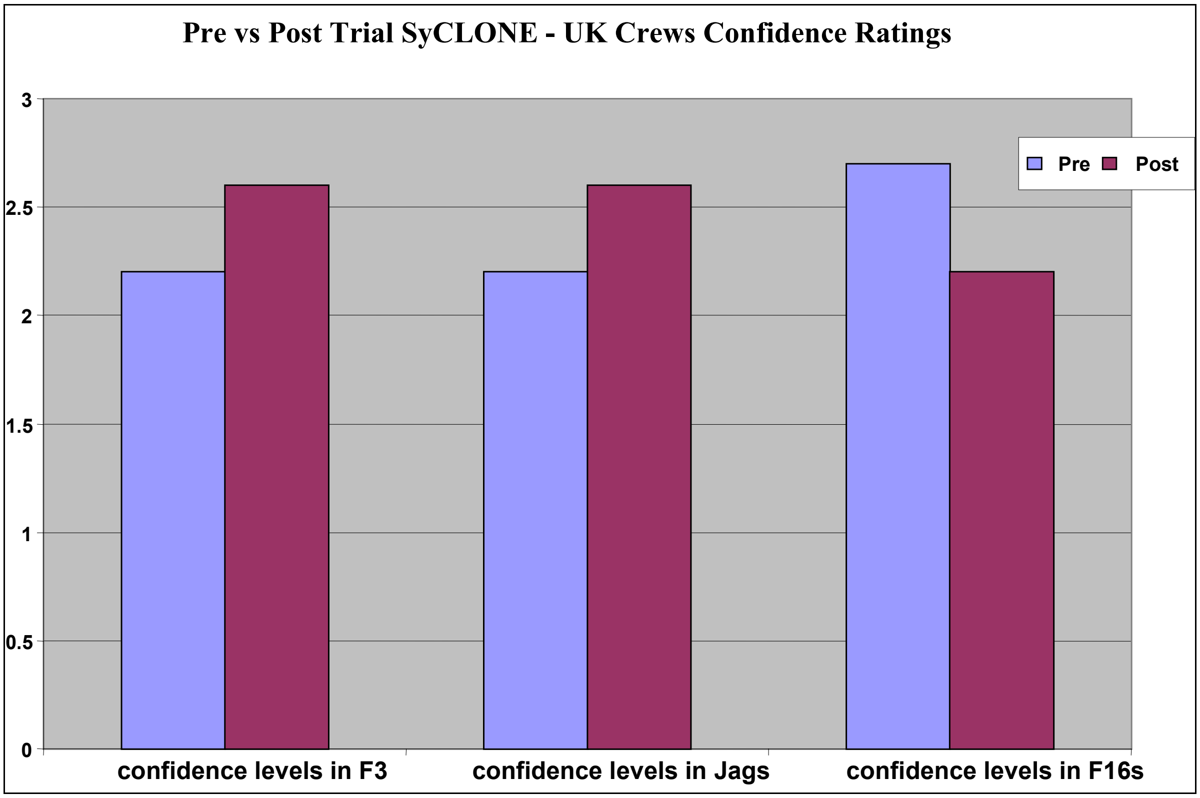


Figure 5: Pre v Post trial UK aircrew ratings for confidence

The results indicted that whilst trust and confidence had increased for the UK participants, it had reduced for the F16 role (the element that was dispersed for the complete duration of the trial). This effect deserves careful consideration.

Trust and confidence between the UK crews may have been able to develop through social interaction, even though attempts were made to keep the A-A and GA formations apart as much as possible. Thus aircrew were sharing the same hotel and took meals together during the trial day. It is also worth bearing in mind that the UK aircrew did manage to have one day colocated where they were able to plan together.

The reputation of the capability of F-16 as weapons platforms may have impacted upon pre-trial ratings. However the lack of social interaction and some of the problems encountered with the PBD technology meant that trust and confidence between the UK and US crews was starting to be eroded during the trial.

Exercise management: Trial SyCLONE provided a first opportunity to test an established, colocated, WF model in a distributed environment; this model has been developed within the research programme and used successfully in previous colocated trials.

SyCLONE has shown that the WF model used in previous, successful, networked trials cannot be applied if participants are distributed. Careful consideration needs to be given on whether or not an alternative exercise management model could be adopted for a real distributed collective training system. One option may be to have WF ‘agents’, or WF liaison officers (WFLOs), embedded with the aircrew. Each WFLO would liaise with the central WF team on the overall gameplan and general thrust of the mission. WFLOs would then monitor aircrew during the planning phase and guide planning direction, where and if appropriate, within the WF

Commander's directive. WFLOs could also pass on relevant information on aircrew performance to the central WF team.

To summarise, running a distributed training exercise is more difficult than running a colocated one. It is not possible to simply apply the model which has been used successfully in colocated trials to the distributed situation. Specifically, an appropriate exercise management model needs to be developed. A central exercise management team cannot have direct oversight of multiple dispersed sites. In order to maintain co-ordination there may be a requirement for WFLOs at each site but this needs to be tried and tested. As yet there is no evidence that the WFLO model would work in practice. However, it is planned to use WFLOs in the NATO SAS34 MTDS exercise, First WAVE, scheduled to take place in September 2004. This exercise should be monitored closely to learn as much as possible about the feasibility of using WFLOs in large scale distributed training exercises.

SyCLONE - overview of findings

There is a limit to the conclusions that can be safely drawn from this trial. SyCLONE was, after all, a single trial based on a small sample of aircrew. However, SyCLONE gave valuable insight into many of the issues associated with running a distributed training exercise. It is a much more difficult prospect than any encountered previously in the research and simply applying the RAPTORS model, known to work successfully in colocated trials, did not work for the distributed situation.

In simple terms the trial can be said to have been a success in that the available technology supported a distributed exercise, in which aircrew planned, briefed, flew and debriefed a complex mission. However the real issue is to understand how effective such an exercise could be in providing high value mission training. The difficulties aircrew experienced might be attributable to distribution, to artefacts of the technology chosen for the trial or simply to lack of experience in using such technology. Some of the observed adverse effects might be overcome by the application of simple expedients. The trial has however clearly demonstrated that it would be a mistake for MTDS protagonists to assume that it is simply a case of 'provide linking technology and all will be well.'

If there is a genuine causal link between poor mission performance and distribution then it is important to understand why this should be the case. One possibility is that the planning constraints already described had an adverse effect on the mission itself. It can be readily understood that these could influence the tactical nature of the plan with undoubted consequences for mission outcome. In order to test this hypothesis it would be necessary to eliminate the known planning constraints and observe the effect on overall performance.

Conclusions

The MoD sponsored research, coupled with military opinion, strongly suggests that there is much benefit in distributed mission training (i.e. MTDS). It has the potential to provide guaranteed COMAO training. The question is, what is an optimised MTDS configuration; all assets dispersed, all colocated or a compromise between the two? MTDS does not necessarily mean that the training systems have to be de-centralised. It is equally possible to conduct training with synthetic training devices all at the same site, linked together via a LAN. The devices would still all be distributed albeit on the same network. However, if the aspiration is to undertaken coalition mission training some elements of MTDS will have to be geographically dispersed. This would add additional levels of complexity due to interoperability issues that may severely impact upon the quality of training that could be undertaken. Research findings on the impact of distribution are not straightforward but a fully distributed MTDS solution would sacrifice some of the intrinsic advantages of colocation.

Irrespective of the final MTDS solution (colocated and/or comprising geographically dispersed elements) the requirement is first fully scope out the problem space in terms of the cost versus training requirement. This necessitates a thorough appreciation of both the technical and training interoperability issues associated with MTDS. It is imperative that there is inter-system and inter-team component compatibility and interoperability. Careful thought must be given to the most appropriate network architecture. In addition there are significant security implications for a military exercise being conducted in a dispersed STE. The optimal technical solution for a distributed mission training capability which fully supports the operational needs of the front-line, yet is cost-effective, and meets all the necessary doctrinal and security requirements has yet to be defined.

References

1. DERA/AS/FMC/CP010229 Smith E & McIntyre H *Simulation and collective training - creating a total experience* The Royal Aeronautical Society, Flight Simulation Group Conference on "Aircrew Training-Time to Take Stock", London, 16-17 May 2001

2. DSTL/CP03983. McIntyre, H & Smith E, Dr Bennett, W (AFRL, Mesa AZ, US) Exploiting High Fidelity Simulation For Aircrew Coalition Mission Training Interservice/Industry Training, Simulation and Education Conference, Orlando, Florida Dec 3rd- 5th 2002

3. Guidance about conducting collaborative working in a distributed command environment. DERA/KIS/SEB/TR001230/1.0, November 2000

4. Guzzo, R.A. and Shea, G.P. (1992). Group performance and inter-group relations in organisations. In M.D. Dunnette and L.D. Hough, (Eds.): Handbook of Industrial and Organisational Psychology

5. DERA/AS/FMC/CP000396 McIntyre, H & Smith E Training In A Synthetic Environment For Improved Operational Effectiveness In Collective Air Operations Interservice/Industry Training, Simulation and Education Conference, Orlando, Florida Nov 27th –30th 2000

Aircrew Mission Training via Distributed Simulation (MTDS) – Development of the Multi-Country Complex Synthetic Environment

David A. Greschke
Simulation Technologies, Inc
Air Force Research Laboratory Bldg 558
Warfighter Training Research Division
6030 South Kent Street
Mesa, Arizona 85212-6061
USA
+1 480 988 6561 ext 429 (Tel.)
+1 480 988 6560 (Fax.)
david.greschke@mesa.afmc.af.mil

Dr. Stefano Cerutti
Alenia Aeronautica S.p.A.
LSIT – Building 5 – Box 18
Corso Marche, 41
10146 Torino
ITALY
+39 011 756 2473 (Tel.)
+39 011 756 2517 (Fax.)
sc@aeronautica.alenia.it

Recent military conflicts have been predominantly multinational in their execution. It has also been an environment involving the use of high technology systems against relatively rudimentary and widely dispersed enemy forces. The lack of well-defined enemy lines and infrastructures has mandated aerospace operations of an increasingly dynamic nature. In addition, due to the "CNN effect" the world is demanding minimal collateral damage. Because these missions involve a complex system of systems, numerous individuals and teams must interact with one another in order to plan and execute each mission. The training of the aerospace forces in this time-compressed, highly dynamic battlespace thus presents a major training challenge.

While teams may have developed high levels of expertise in team tasks within their particular specialties, cultures and nationalities, they may not have similar levels of expertise involving the inter-team tasks needed to effectively operate as part of a larger dynamic aerospace force. Furthermore, all teams at a national level have limited opportunities to train together in realistic, collective training environments.

One potential cost-effective solution to this training challenge is the use of Distributed Mission Training technologies to create a distributed virtual training environment in which the various teams and nations can collectively train together from home station in a common virtual environment.

This paper describes the technological aspects of Exercise First WAVE, a coalition force, Composite Air Operations (COMAO) mission training exercise using distributed simulation technologies being sponsored by the NATO SAS Panel Task Group 034. During this exercise, British, Canadian, Dutch, French, German, Italian and US aircrews, together with command and control personnel will plan, brief, execute, replay, and debrief composite force air missions using real-time simulators located in their home countries and interconnected via secure data links. The objectives of the project are to expand distributed simulation to intercontinental distances; develop systems to mitigate difficulties caused by extreme long distance links; establish processes for creating scenarios to fulfill specified training objectives; implement systems for distributed mission planning, briefing, and debriefing; and assess the effectiveness of distributed simulation for enhancing Warfighter skills in conducting coalition force operations.

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held in Genoa, Italy, 13-15 October 2003, and published in RTO-MP-HFM-101.

Keywords:

Mission Training, Distributed Simulation, Networked Training, Collective Training, Coalition Training, Complex Synthetic Environments, Virtual Simulation, Transoceanic Networks

1.0 INTRODUCTION

Mission Training via Distributed Simulation (MTDS) refers to several simulation resources, distributed over a geographically extended area, linked by a common network infrastructure. The primary use for such a distributed system as described in this paper is to conduct Composite Air Operations (COMAO), for training and/or mission rehearsal purposes. Of course the operations need not be restricted to the aeronautics domain: any type of resource can populate the distributed Synthetic Environment, thus including surface, sub-surface, ground and space entities as well.

Exercise First WAVE (EFW – Warfighter Alliance in a Virtual Environment) has been designed with the intent of demonstrating a NATO multi-national MTDS exercise, using HLA as the underlying architecture, the HLA FEDEP as a guiding development process, and existing legacy simulators as main actors, along with computer-generated forces (CGF). The entire project is being developed under the auspices of NATO through its Research and Technology Office and the Studies, Analysis and Simulation (SAS) Panel and NATO Modeling and Simulation Group (MSG), as shown in figure 1. The MTDS demonstration, First WAVE, is the deliverable for the joint task group effort.

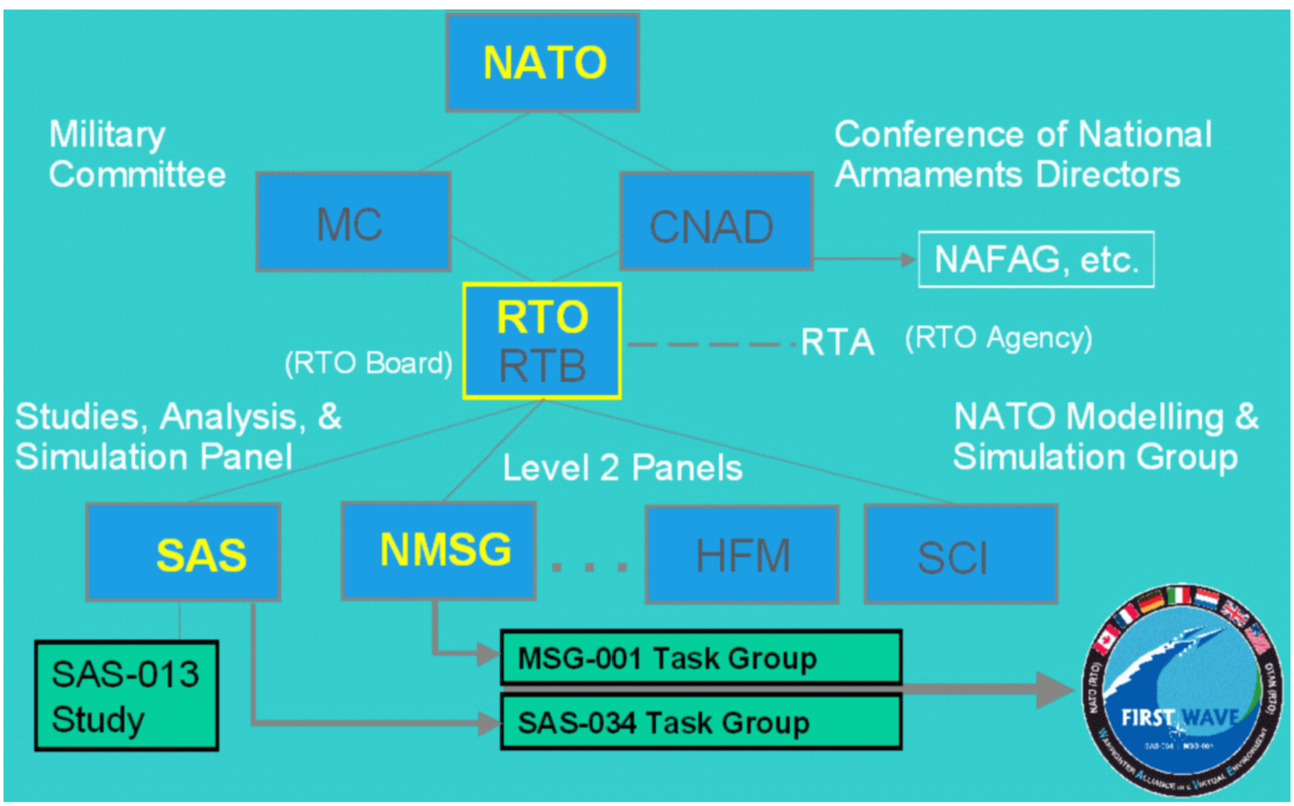


Figure 1: Position of the Task Groups within the NATO Structure.

Assessment of the feasibility, repeatability, and training value of NATO MTDS will be part of the final result, available some time after the demonstration which is currently scheduled for the third week of September 2004. Additional analysis will be devoted to the exploration and resolution of security concerns and challenges (the classification will be NATO Secret releasable to the participating nations only, namely Canada, France, Germany, Italy, The Netherlands, United Kingdom, and the United States), exercise management, and the performance and realistic behavior of CGFs as perceived by the pilots.

This paper describes the basic components of EFW and the technical challenges that have been tackled so far. Since it is still a "work in progress", some of the information herein presented is subject to change before the final demonstration. For a less technical overview of the same project, which also takes into account previous studies and NATO activities (such as SAS-013), please refer to "Aircrew Mission Training via Distributed Simulation – Progress in NATO", session V, paper number 16, by B. Tomlinson and J. van Geest.

2.0 THE EXERCISE

During the first 15 months of its existence, the First WAVE group operated as a main task group (Steering Committee) with five task teams (see figure 2). Each of the Task Teams has its own, domain-specific objectives, and documents have been produced to provide more detail regarding the Technical, Security and Assessment Teams objectives.

- **SAS-034 Task Group:**
    - Chairman, Barry Tomlinson (QinetiQ UK)
- **The five Task Teams within the Task Group:**
    - **Operations & Training Task Team**
        - Chairman: Capt Lee Gibbard (DND CA)
    - **Technical Task Team**
        - Chairman: Dave Greschke (AFRL US)
    - **Security Task Team**
        - Chairman: Lt Col Bob Jones (DND CA)
    - **Assessment Task Team**
        - Chairman: Dr. Herb Bell (AFRL US)
    - **Public Relations/Awareness Task Team**
        - Chairman: Barry Tomlinson (QinetiQ UK)

Figure 2: The SAS-034 Task Team Structure.

In September 2002 the United States announced it would fund the hiring of a First WAVE Project Manager to oversee the "big picture" and the project management plan which had been written. The key to success will be the test and integration plan, which was originally planned for a 12 to 18 month period. Communication costs, and the momentum experienced in most government budgeting processes, has led to the point that the real testing plan will be nine months for some countries and sites, and six months for others. These conditions are about the minimal time frame that can be reasonably tolerated without jeopardizing EFW, scheduled to run in September 2004.

In this section, the objectives of EFW are described, and the exercise mission illustrated.

2.1 The Objectives

The overall objectives of exercise First WAVE are:

- To investigate and assess the potential of MTDS to enhance aircrew mission readiness for NATO coalition operations, including consideration of training, simulation technology and management aspects
- To increase awareness of MTDS capabilities amongst national and NATO military staffs
- To capture experiences and lessons that will influence the future direction of the NATO MTDS program.

In addition to these overall objectives, there are specific national objectives. It will be the goal of the MTDS design for Exercise First WAVE to satisfy objectives at a multi-national or “alliance” level. However, while there is consensus of opinion in the overarching exercise objectives, it is anticipated that training and research objectives may vary according to the different national imperatives into synthetic mission training. National representatives of the SAS-034 and MSG-001 Task Groups may, therefore, wish to assess the Exercise according to their own National training and research criteria. The latter assessment processes are not summarized in this paper.

The overall objectives have been broken down into Training Objectives and Technical Objectives, which translate into the corresponding requirements. To the maximum extent possible, the following training objectives will be used as a basis of assessing the training utility of the MTDS demonstration:

- To practice daytime COMAO procedures employing fighter escort/sweep, AAR, SEAD, RECCE and AEW in a hostile EW environment
- To exercise procedures for defensive operations with Fighter Areas of Responsibility (FAORs) and point defense tasking
- To employ EW resources in support of offensive and defensive air operations
- To plan and integrate a multi-national COMAO in a defined threat environment
- To brief a COMAO package generated from dispersed locations
- To conduct mission debriefs
- To engender efficient team-working skills between Nations and differing elements of the COMAO package
- To develop a tactical appreciation of real-world threats
- To expose aircrew to situations to which they would not normally encounter in a peacetime environment
- To establish lessons identified.

Exercise First WAVE will also be used to facilitate investigation into a number of key areas.

1) Exercise Management, focusing on scenario development, process and strategy, tools, monitoring, control and analysis of mission critical events, and management of a distributed coalition exercise involving multiple sites
2) Interoperability, focusing on database requirements, utility of secure WAN, security policy, procedures, systems, federation design and integration process, standards, tools, voice communications, database exchange and re-use
3) Computer Generated Forces, focusing on their credibility, control, portability, and flexibility.

The technical objectives which had to be considered throughout the life of the project can be summarized as follows:

- Expand distributed simulation to intercontinental distances
- Develop systems, devices and procedures to mitigate difficulties caused by extreme long distance links
- Establish processes for creating scenarios that fulfill specified training objectives
- Implement systems for distributed mission planning, briefing, and debriefing
- Assess the effectiveness of distributed simulation with respect to enhancement of Warfighter skills in conducting coalition force operations.

It was recognized and agreed upon early in the planning that the First WAVE infrastructure would consist of devices of varying fidelity. The major reason was that there is no external source of funding other than each nation's contribution, so it was agreed that legacy simulators were going to be accepted "as is". As a consequence, a mix of full field of regard, high fidelity cockpit simulators will have to interoperate with single channel visual part task trainers, for instance. This is not the most ideal situation, but corresponds to the real world situation which needs to be taken into account. After reaching this agreement, the scenario was designed to accommodate the capabilities of the simulators used by each country. The only strictly necessary requirement was that all simulators participating had to be capable of maintaining real time. It must be finally pointed out that the proof that legacy resources can be successfully reused in a distributed simulation is an added value in itself.

2.2 The Mission

Air missions in a NATO context now focus extensively on operations where 20 to 40 or more aircraft fly in a package to strike a specific target or a set of targets. The composition of any package is based on the type of target, the expected threat during the mission and the level of destruction desired on the target.

Such a mission is referred to as a Composite Air Operation. COMAOs are defined as "operations interrelated and/or limited in both time-scale and space, where units differing in type and/or role are put under the control of a single commander to achieve a common, specific objective". In this context, some typical roles are: Strike Attack, Air Defense, Offensive Support and Reconnaissance. Support roles include Airborne Early Warning, Electronic Warfare and Air-Air Refueling. A COMAO package will comprise aircraft in defined formations performing specified roles. A specific COMAO may also be referred to as a mission. Successful participation in a COMAO requires that aircrews be prepared to participate as effective members of a multi-national force: these aircrews must be trained to operate as part of a collective combined force involving two or more teams from two or more countries. To meet this requirement, aircrews must master the skills necessary not only to employ their individual weapons systems but also master a number of collective or inter-team skills involving communication, co-ordination, planning, decision making, and situation assessment that will be exercised in a complex multinational environment.

COMAO training focuses on collective skills, and collective training is therefore defined as "training involving two or more teams, where each team fulfils different roles, training (to interoperate) in an environment defined by a common set of training objectives".

Exercise First WAVE has been designed to demonstrate the potential of MTDS throughout an end-to-end COMAO involving the following four phases: planning (including brainstorming), briefing, flying and debriefing. Activities designed to provide planning, briefing and debriefing activities are discussed below, while an illustration of the strike package air operations is summarily presented later.

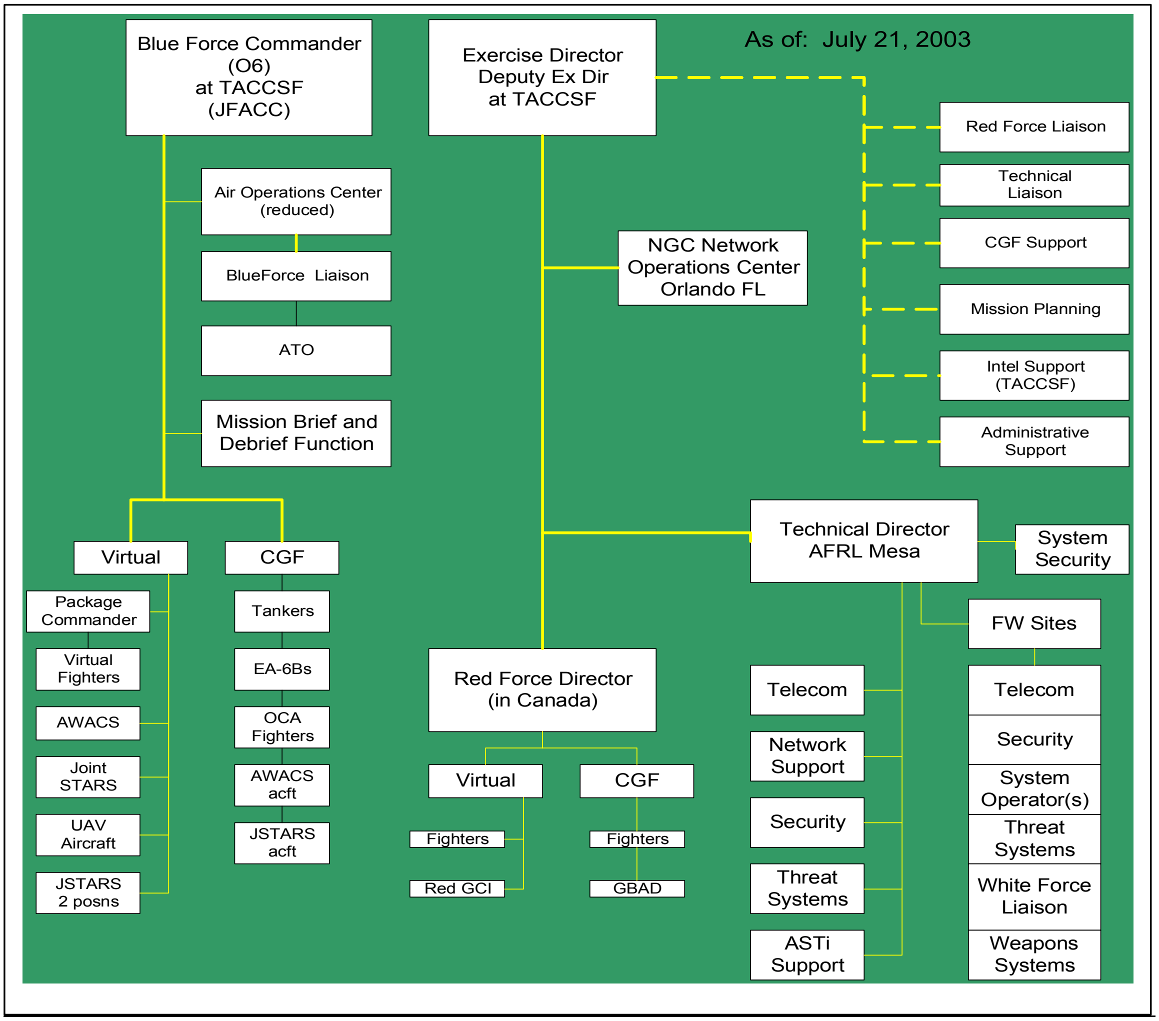


Figure 3: Functional Layout of the Overall First WAVE Execution Management.

A Typical COMAO mission starts when the Air Tasking Order (ATO) is first received. The ATO provides some very detailed information including Mission aim, package composition (i.e. the number and type of aircraft), weapons loads, target details etc. The ATO also specifies who the Mission Commander (MC) will be. Although the ATO is normally issued 8 to 10 hours before it is implemented, for Exercise First WAVE the ATO will be issued at the beginning of each mission day. Once the ATO has been 'broken out' by the MC, the MC will form a brainstorming and planning team to plan and co-ordinate the mission, having first received weather ("Met") brief, and an intelligence update.

Planning starts with the analysis of the target (type, location, and desired effects), taking into account the weather and the likely threats to be encountered *en route* and in the target area. The MC, assisted by his planning team, will formulate an attack plan. This plan will give due consideration to the package composition, DMPIs (Designated Mean Points of Impact) weapons, individual TOTs, package flow, ingress and egress routes, code words, and so on. Particular attention will be made to deconfliction plans.

The MC will also co-ordinate all other package elements such as electronic protection, SEAD (Suppression of Enemy Air Defense), fighter protection, and reconnaissance. This will include the overall routing, the overall package flow, the fighter sweeps and CAP (Combat Air Patrol), the RAI (Reconnaissance - Attack Interface), the AAR (Air-to-Air Refueling) plan, the AWACS plan, pre-strike holding and so on.

The Air Tasking Orders will be produced in advance by the Mission Planner. The blue force aircrews will be given additional information (intelligence, safety, weather, SPINS) to enable them to plan and brief. Facilities and equipment should be available to the aircrew at all nodes, to enable them to plan and brief.

For debrief, the ability to fully reconstruct the mission, enabling replay of any mission event, is desired. However, debrief capabilities will be a local responsibility at each of the nodes. The debrief will be led by the Mission Commander and must be accessible to MTDS participants at all sites. A synchronized mission play-back capability is planned for each site. Use of SmartBoard and VTC equipment is desired at as many sites as is possible in order to promote aircrew and White Force interaction. This will also facilitate a more effective debriefing and lessons learned session.

The map in figure 4 (yellow dashed outline) shows the gaming area, while figure 5 shows an overall view of the battle plan with friendly forces based in Italy and red air and ground forces scattered throughout the game area. Air refueling tracks and rendezvous airspace have been designated as shown in the same figure; the north and south edges of the battle area contain the orbits for high value assets such as the Joint STARS and AWACS aircraft. The blue forces comprise a mix of boom and hose/drogue air refueling tankers, as can be seen from figure 6. Blue Forces will be based in and will take off from four different bases in Italy, the only exception being a pre-positioned Forward Air Controller operated by The Netherlands, which will be in the target area on the ground at time zero. The First WAVE Strike Package configuration following the air refueling and rendezvous phases is shown in figure 8. The mission will unfold from this point on, based on real-time actions and interactions, according to their own flight plans. An example of ground targets is shown in figure 9 (targets 13, 14, 15 and 16), consisting of fuel storage tanks. Figure 10 shows targets 17 through 20, which are a series of building complexes.

The approximate locations of on-station assets are shown in figure 7.

Figure 4: First WAVE Gaming Area (yellow dashed box).

Figure 5: Overall View of the Battle Plan.

Air-Air Refueling TANKERS - BOOM
- 2 x KC-10 (*KILO CHARLIE 12)*
- 2 x KC-10 (*KILO DELTA 23)*
- 1 x KC-10 (*KILO GOLF 14*)

Air-Air Refueling TANKERS - HOSE/DROGUE
- 2 x KC-10 (*KILO ECHO 33)*
- 2 x KC-10 (*KILO FOXTROT 25)*
- 1 x KC-10 (*KILO HOTEL 15*)

BRINDISI / CASALE AIRBASE
- F-15C (*STEIN 61*) TACCSF
- F-15C (*GHOST 71*) EGLIN
- F-16C (*VIPER 41*) MESA

LECCE / GALATINA AIRBASE
- UAV (*SPECTRUM 05*) TACCSF
- EFA (*CANNON 31*) GE / IT
- M2000C (*VOXAN 11*) FR

GIOIA DEL COLLE
- GR4 (*SILVER 21*) UK
- GR4 (SAF)
- CF-18A (*COBRA 33*) CA
- F-16MLU (*LOTUS 50*) NL

BARI / PALESE
- F-16CJ (*ADAMS 45)* SHAW
- EA-6B (*COMANCHE 77*)

IN Pristina TARGET AREA
- FAC (*WINDMILL*) NL

Note: Red entries are CGFs

Figure 6: List of the Blue Forces.

AWACS / JSTARS in ORBITS

BOSNIA
- E-3C (*BADGER 01)* TACCSF
- F-15C (*GHOST 13*)

ALBANIA
- E-3C (*DISCO 02*)
- E-8 JSTARS (*BIGTOE 35)* TACCSF
- M2000C (*VOXAN 35*)

AAR TRACKS

NORTH
- KC-10 / BOOM (*KG 14*)

CENTRE
- KC-10 / BOOM (*KC 12*)
- KC-10 / BOOM (*KD 23*)
- KC-10 / HD (*KE 33*)
- KC-10 / HD (*KF 25*)

SOUTH
- KC-10 /HD (*KH 15*)

Note: Red entries are CGFs

Figure 7: On-Station Assets.

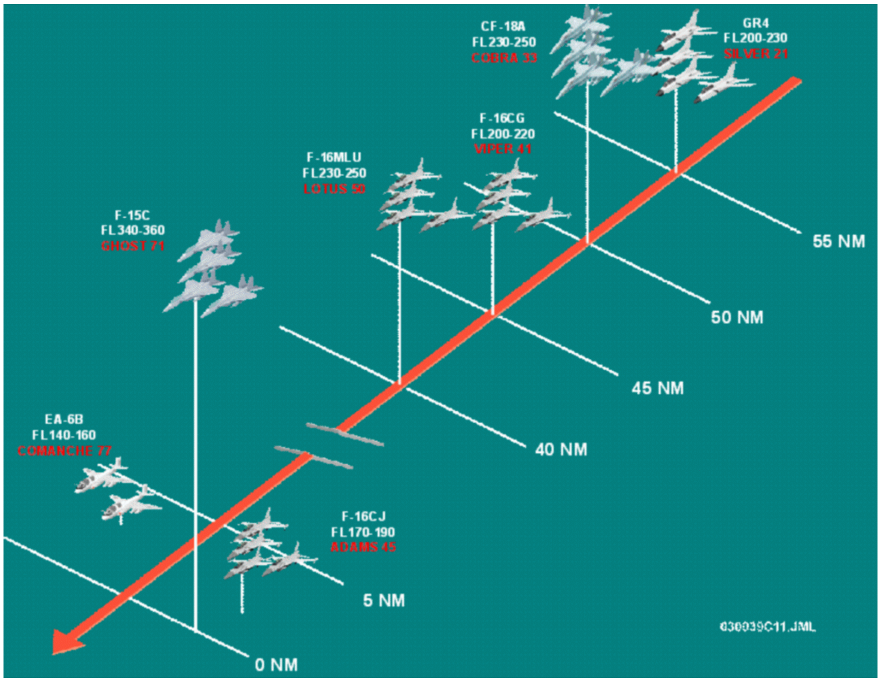


Figure 8: Strike Package Configuration.

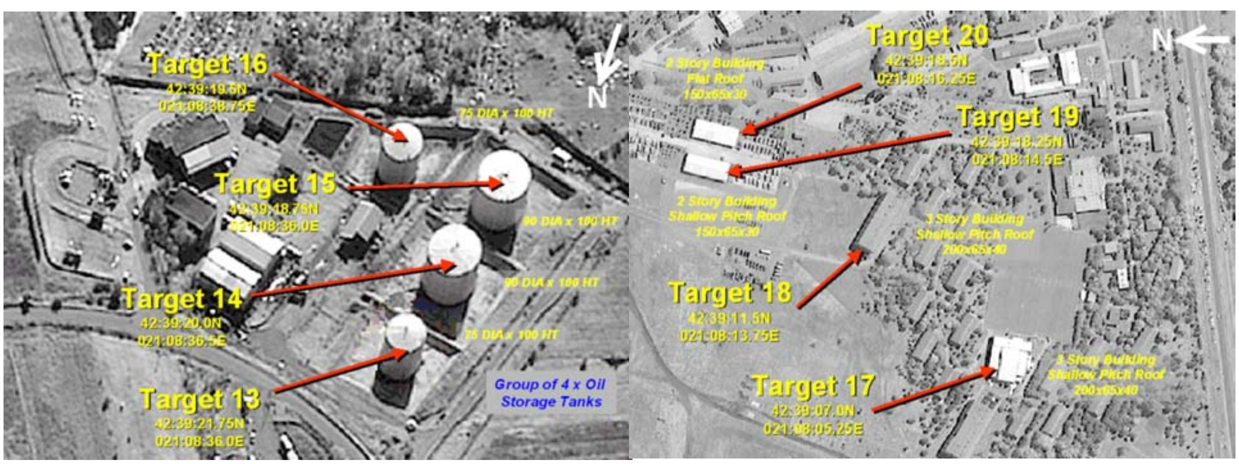


Figure 9 and 10: Example of Target Sets in the High Resolution Area.

Originally a traditional approach to the White Force was anticipated, with all of the required functions located in one place. The plan was revised when Canada requested to manage the entire Red Force and treat it as a blue force exercise from their point of view. Under these conditions, it was agreed to move the Red Force Director function to Canada and move the remaining functions to the Theater Air Command and Control Simulation Facility (TACCSF), located at Kirtland AFB New Mexico (USA). This change was accepted since it maximizes the training benefits for everyone. The only requirement of the "Red Force" execution is to remain subordinate to the real Blue Force Training objectives during First WAVE. The diagram in figure 3 shows the functional layout of the White Force and the overall First WAVE execution management.

The enemy Integrated Air Defense system will be structured as shown in figure 11, its key features being several air bases, fixed and mobile radar and missile sites. The specific force composition includes MiG-29Bs, MiG-21 2000 and MiG-21bis, SA-3 and SA-6. Five enemy air bases will be under the control of the Red Force Director.

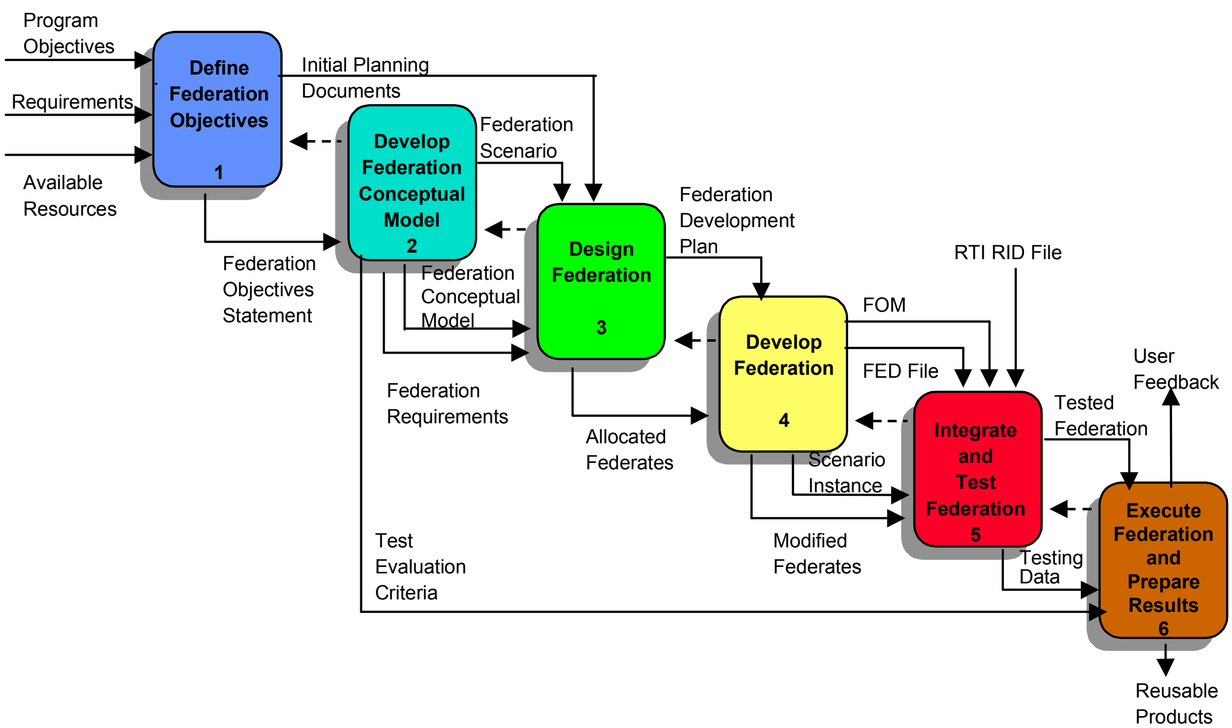
Figure 11: "Notional FW Enemy" IADS.

2.3 The Federation Execution and Development Process

The connection between all simulation resources adheres to the High Level Architecture standard (HLA), which will be described in section 3.2. This architecture specifies a development and execution process called the FEDEP which has been used as the basis for planning throughout the First WAVE project.

When the initial definition of the High Level Architecture (HLA) was first made public in early 1995, a number of new concepts were introduced. One such concept was the notion of a federation, which was defined as a set of software applications capable of exchanging information based on an agreed upon interchange document known as a Federation Object Model (FOM) and brokered by a common federate communication interface referred to as the Runtime Infrastructure (RTI).

One of the design goals identified early in the development of the HLA was the need for a high degree of flexibility in the process by which HLA applications could be composed to achieve the objectives of particular applications. Because of this basic desire to avoid mandating unnecessary constraints on how HLA applications are constructed, it was recognized that the actual process used to develop and execute HLA federations could vary significantly within or across different user applications. For instance, the types and sequence of low-level activities required to develop analysis-oriented federations is likely to be quite different from those required to develop distributed training exercises. However, at a more abstract level, it is possible to identify a sequence of six very basic steps that all HLA federations will need to follow to develop and execute their federations: they are shown in figure 12, and briefly described below.



Step 1: Define Federation Objectives. The federation user and federation development team define and agree on a set of objectives and document what must be accomplished to achieve those objectives.

Step 2: Develop Federation Conceptual Model. Based on the characteristics of the problem space, an appropriate representation of the real world domain is developed.

Step 3: Design Federation. Federation participants (federates) are determined, and required functionalities are allocated to the federates.

Step 4: Develop Federation. The Federation Object Model (FOM) is developed, federate agreements on consistent databases/algorithms are established, and modifications to federates are implemented (as required).

Step 5: Integrate and Test Federation. All necessary federation implementation activities are performed, and testing is conducted to ensure that interoperability requirements are being met.

Step 6: Execute Federation and Prepare Results. The federation is executed, outputs are generated, and results are provided.

Since this six-step process can be implemented in many different ways depending on the nature of the application, it follows that the time and effort required to build an HLA federation can also vary significantly. For instance, it may take a federation development team several weeks to fully define the real world domain of interest for very large, complex applications. In smaller, relatively simple applications, the same activity could potentially be conducted in a day or less. Differences in the degree of formality desired in the development process can also lead to varying requirements for federation resources. It is the hope of the developers of First WAVE that the many hurdles encountered will be considerably easier to overcome if there is a "Second WAVE" and that it will be of benefit to capitalize on the results here achieved and left in place by First WAVE.

3.0 THE VIRTUAL ENVIRONMENT FOR FIRST WAVE

A significant technical challenge that appears whenever many different sites or several different versions of image generators are linked together to create a virtual environment is the problem of database correlation. The First WAVE common exercise area covers 48 geocells (one degree latitude by one degree longitude each), or about 173,000 square miles. The location of the gaming area for First WAVE, Kosovo, was chosen simply based on the fact that the database provided a manageable size for a database, the data for this region was readily available, and there was previous operational coalition experience which could possibly be compared to the virtual recreation of these previous military operations. The yellow outline on the map in figure 4 shows the overall extent of the database for First WAVE. It is rendered overall at 10-meter resolution. The targets sets (figures 9 and 10), all located within a 20-mile radius circle, are generally one-meter resolution insets.

The various sites employ, within their simulators, image generators manufactured by different vendors, with varying degrees of sophistication. Canada volunteered to produce the database in several different formats, compatible with the toolsets of the major image generators used. The formats provided were either Open Flight, for more modern image generators, or the entire collection of Geotiff and DTED files for those sites which had specific format requirements. Another US First WAVE industrial partner agreed to convert the First WAVE database to E&S formats for all of the existing E&S customers. Due to limited budget and engineering integration time, sensor databases and Link-16 implementation over the wide area network were eliminated from the requirement list.

In order to produce the computer-generated forces (CGF) necessary to support First WAVE, it was originally thought to use Joint SAF. Unfortunately it was quickly discovered that Joint SAF is not for export, and an alternative CGF environment had to be identified. A Canadian industrial partner volunteered to support First WAVE with a commercial product, for both generating the CGF environment during First WAVE and to support the development and integration of the Test Federate.

Figure 13 shows a diagram of what a typical laser-guided bomb attack might look like, shown here as an example of the detailed planning that had to go into designing the actual scenario to be used, to ensure that the scenario would be tactically relevant.

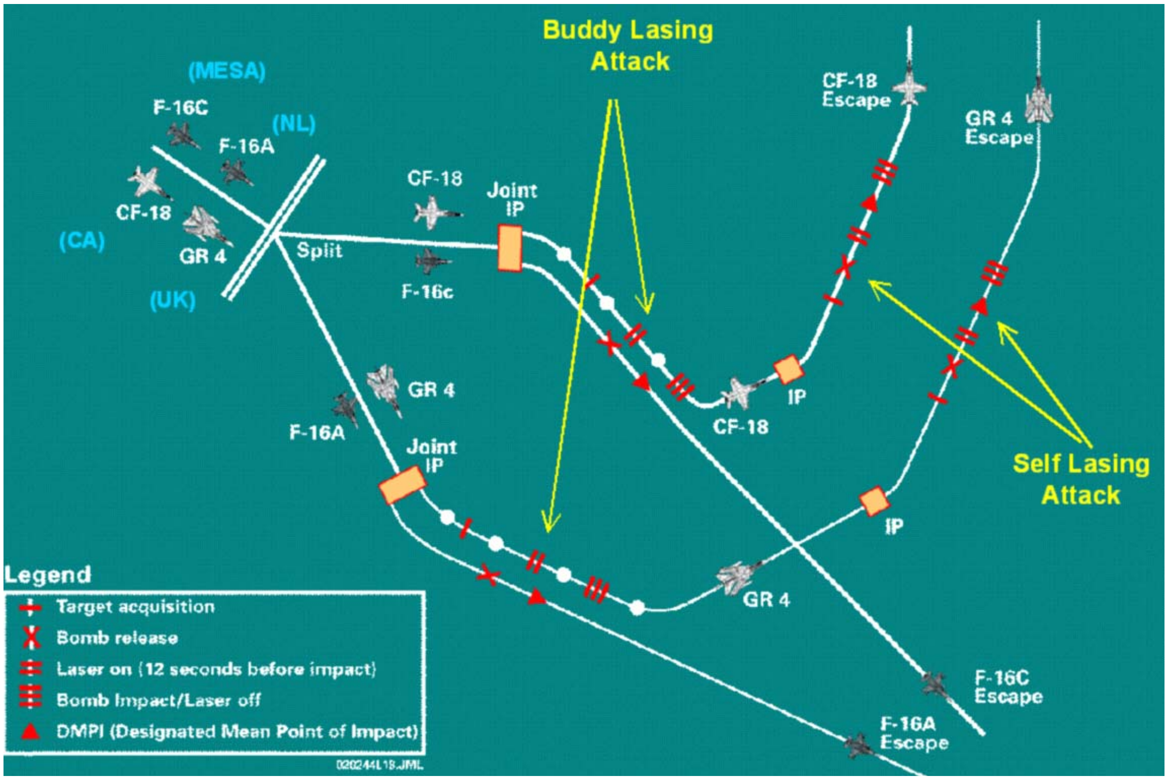


Figure 13: Sample Laser-Guided Bomb Attack.

During the definition phase of the technical objectives, the determination of the extent of the network itself, figuring out its physical size and the number of entities in the scenario were all critical factors that had to be taken into consideration in order to determine the best approach for the requirements on the physical network. This forced a tight cooperation among the Task Teams within First WAVE. On top of the physical network, software architecture had to be chosen, and the entire system will need to be accredited by the National Security Agencies in order to operate at the required classification level.

Presently, the seven countries that will participate in First WAVE are Canada, France, Germany, Italy, The Netherlands, United Kingdom, and the United States. The network will consist of 15 separate locations where manned simulators are present, in addition to the Network Operations Center (NOC) located in Orlando, Florida.

3.1 The Network

Issues of packet latency, capacity, quality of service and reliability had to be investigated. Most sites will also include a recording and playback software and hardware package called DCS (DMT Control Station), and provided free of charge by the Air Force Research Lab in Mesa, AZ. Voice communication will be taken care by the ASTi radio simulation environment during the actual flying time, and by specific videoconferencing equipment for briefing and debriefing.

Due to the physical size of the network and the predicted number of entities involved, a full mesh ATM backbone was chosen for First WAVE. The network activity and performance will be controlled by the

Network Operations Center (NOC) in Orlando, Florida. The network scheme (figure 14) includes what is known as the DMT Portal, which is a PC-based layer above and below an encryption device. The portal is the common interface among all sites, acting therefore as a gateway, whose configuration can only be controlled by the NOC.

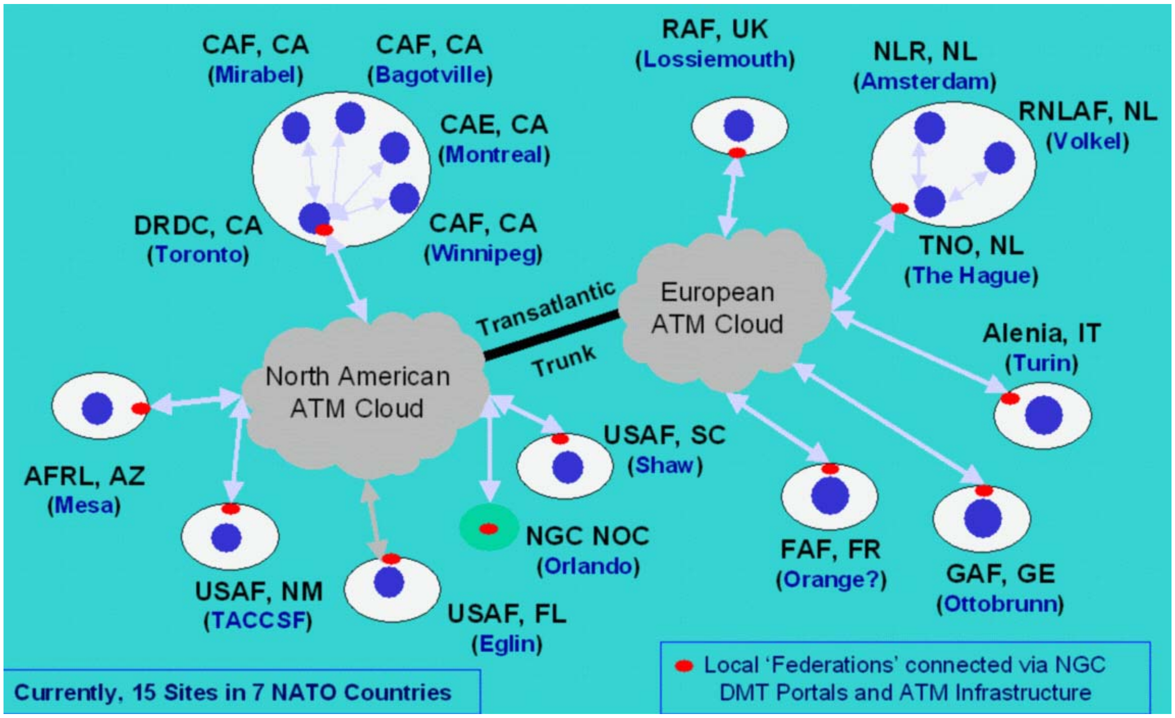


Figure 14: The Network Topology.

3.2 The Architecture

Several protocols can be used to shuffle data among networked sites. With a complex scenario, though, what really makes the difference between success and failure is a suitable architecture. Among all of the decisions that had to be made early in the process, one of the major ones was which interoperable protocol architecture to use. Even though the overall desire was to use High Level Architecture (HLA), it was decided to allow the use of both HLA and Distributed Interoperable Simulation (DIS). The overall categorization of First WAVE is an HLA exercise; however, since there was no external budget, those sites running DIS are allowed to participate. The sites with a native HLA solution are allowed to run HLA, and the DMT Portal will handle either case (HLA and DIS) and manage the communications over the long haul, thereby reducing the integration time needed for interoperability.

Most legacy flight simulators have been designed for stand-alone operation, principally aimed at skills development and maintenance. Typical MTDS applications require, on the contrary, a substantial level of interoperability and training at the team and teams-of-teams level. As a consequence, a considerable amount of optimization and integration is normally needed. The High Level Architecture was designed for such a project but there a still very few real-time networked simulators operating with a native HLA solution. Thus, First WAVE will support both protocols.

Within virtually all western nations, in particular the USA and the UK, there is a substantial effort to integrate and acquire HLA know-how, and to use it as a replacement for earlier DIS protocol. First WAVE will serve as a high fidelity, large scale proof-of-concept for fast jet real-time environments.

3.3 Security

From the very beginning, the Technical Task Team and the Operations and Training Team insisted that First WAVE be conducted at the SECRET level. The reason for this requirement was to maximize the training value of the entire project. A specifically designated sub-group began investigating whether or not there were any existing agreements already in place between participating nations that would make the approval process easier: none suitable were found.

The one existing agreement and corresponding network that looked promising for First WAVE is known as the Combined Forces Battle Lab (CFBL). The reason CFBL was not chosen is that some of the networks it comprises have too little bandwidth to support all the traffic that will be generated, in addition to the fact that First WAVE would not have had the priority it needs in 2004, due to JWID and JEFX, two major battlestaff exercises scheduled for mid-2004.

Current efforts are being made towards gaining accreditation from the Multinational Security Accreditation Board (MSAB). The current MSAB members are the USA, Canada, and the United Kingdom. Details are being worked out to include The Netherlands, Germany, France and Italy, and it seems that the final stamp of approval will be issued swiftly, once each nation approves the exercise according to its own security accreditation criteria.

The current classification of First WAVE remains “NATO SECRET – Releasable to CA, FR, GE, IT, NL, UK, US”, and a joint Memorandum of Understanding (MoU) and Security Agreement is being worked on at this time.

4.0 CONCLUSIONS

The key benefits of First WAVE are several. Its primary characteristic is that it will most likely be the largest real-time, distributed, virtual exercise ever attempted. From Arizona to Italy, it will be working with probably the largest number of real-time HLA federates ever connected. This will enable testing of HLA in a very unique and demanding environment – fast jets within visual range of each other. The large area network will also provide excellent insight into the characteristics of transoceanic networks with regards to latency and stability issues.

First WAVE will serve as an excellent proof of concept and validation of global distributed mission operations in a training-on-demand coalition environment. It will leave behind valuable knowledge, a multinational training infrastructure, and will establish an important milestone in how to approach and hopefully solve the problems that need to be dealt with in such a complex endeavor.

5.0 LIST OF ACRONYMS

AFRL	Air Force Research Laboratory
CEPA	Common European Priority Area
CGA	Computer Generated Actor
CGF	Computer Generated Forces
CNAD	Conference of National Armament Directors
COTS	Commercial Off The Shelf
DERA	Defense Evaluation and Research Agency
DIS	Distributed Interactive Simulation
DMSO	Defense Modeling and Simulation Office
DOD	Department Of Defense
DTED	Digital Terrain Elevation Data
DFAD	Digital Feature Analysis Data
FLIR	Forward Looking Infra-Red
FMS	Full Mission Simulator
FOR	Field Of Regard
HLA	High Level Architecture
HMD	Helmet/Head Mounted Display
HMI	Human Machine Interface
HUD	Head-up Display
HW	Hardware
IEEE	Institute of Electrical and Electronic Engineers
IR	Infrared
ISDN	Integrated Services Digital Network
LMA	Locally Manned Aircraft
MC	Military Committee
MIT	Massachusetts Institute of Technology
MTDS	Mission aircrew Training via Distributed Simulation
NVG	Night Vision Goggles
PVI	Pilot Vehicle Interface

RPV	Remotely Piloted Vehicle
RTA	RTO Agency
RTB	RTO Board
RTI	Run-Time Infrastructure
RTO	Research and Technology Organization
RTP	Research and Technology Project
SA	Situation Awareness
SAS	Studies, Analysis and Simulation Panel
SBA	Synthetic Based Acquisition
SE	Synthetic Environment
STEP	Scenario for Test and Evaluation Purposes
STOW	Synthetic Theatre Of War
SW	Software
TRA	Training, Rehearsal and Acquisition
UAV	Uninhabited Aerial Vehicle
UCAV	Uninhabited Combat Aerial Vehicle
USAF	United States Air Force
WEAG	Western European Armament Group
WEU	Western European Union
WRC	WEAG Research Cell

Adaptive Thinking Training For Tactical Leaders

Dr. James W. Lussier
U.S. Army Research Institute - Fort Knox
Armored Forces Research Unit
Bldg 2423 Morande Street
Fort Knox, KY 40121-5620
USA
Tel: 502-624-2613 Fax: 502-624-8113
E-Mail: James.Lussier@knox.army.mil

Mr. Scott B. Shadrick
U.S.Army Research Institute – Fort Knox
Armored Forces Research Unit
Bldg 2423 Morande Street
Fort Knox, KY 40121-5620
USA
Tel: 502-624-2613 Fax: 502-624-8113
E-Mail: Scott.Shadrick@knox.army.mil

Summary

This paper reports a series of research efforts embodied in the U.S. Army's Think Like a Commander training program. The work is interesting because it seeks to train a cognitive behavior – thinking – using methods that have traditionally been applied to training more observable and measurable behaviors, e.g., marksmanship and gunnery, sports performance. In short, it does not greatly respect a traditional distinction between such things as physical movements, perceptions, and cognitions when it comes to training, rather treats these all as behaviors that are amenable to the same training methods and principles. Deliberate practice techniques were applied to develop exercises to train the task of adaptive thinking in tactical situations. The exercises were used in command preparation courses for U.S. Army officers at the brigade, battalion, and company levels. The approach shows promise and initial data indicate significant performance gains in a key component of battlefield adaptive thinking: the rapid analysis of battlefield situations to identify of key considerations for decision-making.

Soviet Chess Training Methods

The notion that thinking can be trained as a behavior has a precedent in practice. For decades, the Soviet chess machine thoroughly dominated all competition. Chessplayers around the world assumed the Soviets achieved their success solely by extra effort in selecting, developing, and supporting promising players. But did the Soviets have some new and secret training methods that the rest of the world did not? Few imagined that. With the breakup of the USSR, Soviet chess academies became publishing houses. The release of such books as Mark Dvoretsky's *Secrets of Chess Training* and *Positional Play* surprised the chess world. It seemed that the Soviets did have methods they hadn't revealed. Subsequently, English-speaking chess trainers have written manuals that applied the Soviet methods to selected aspects of the game, for example Andrew Soltis' book on how to calculate in chess, titled *The Inner Game of Chess*, presents training exercises to develop skill at visualizing future positions by moving pieces in one's imagination.

Researchers at the U.S Army Research Institute (ARI) saw a parallel between the problem of training battlefield commanders to think adaptively in tactical situations and that of training chess grandmasters. They analyzed the Soviet training manuals to understand their methods. The difference between the Soviet methods and traditional chess instruction is, in a sense, the difference between education and training. The rest of the world studied the game of chess, its strategies and tactics, and tried to understand why one move was better than another. As students studied the game, they acquired knowledge about chess and understanding of its principles. They educated themselves about the game of chess. The Soviets did that as well, but also studied the human processes of finding good moves and avoiding errors, of searching and evaluating chess positions, and of controlling emotion and fighting a psychological battle with one's opponent. The Soviets described principles of expert play that reflected the thought patterns of grandmasters. While many of these expert principles were familiar to the rest of the world, the Soviet

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

trainers went one critical step further. They created exercises that trained these principles, ingraining them in their students. After sufficient training, the Soviet students employed the expert thought patterns not simply because they understood the principles nor because they were consciously directing their thinking by using the expert patterns as a checklist. The cognitive behaviors had become automatic. As a result of the exercises, the students followed the principles without thinking about them, freeing their limited conscious resources to focus on the novel aspects of the contest and to think more deeply and creatively at the board. The Soviet chess trainers in essence treated the thinking that the player does during a game as a behavior – something a player does with chess knowledge as opposed to the knowledge itself – and then developed exercises to train that thinking performance to conform to that of an expert.

Adaptive Thinking on the Battlefield

The cognitive task to which our research effort has been applied is what has come to be called adaptive thinking by the U.S. Army. The term adaptive thinking, as we define it, "describe[s] the cognitive behavior of an officer who is confronted by unanticipated circumstances during the execution of a planned military operation" (Lussier, Ross, & Mayes, 2000, p.1). The skillful commander will, when performing adaptively, make adjustments within the context of the plan to either exploit the advantage or minimize the harm of the unanticipated event, in short, will adapt to conditions for a more successful outcome. This description of the adaptive thinking task defines the behavior in terms of the problem to be solved – to monitor the unfolding tactical situation for unanticipated events and to determine the proper actions in response to them. Another important aspect of the task involves the conditions under which it must be performed. The thinking that underlies battlefield decisions does not occur in isolation or in a calm reflective environment, it occurs in a very challenging environment. Commanders must think while performing: assessing the situation, scanning for new information, dealing with individuals under stress, monitoring progress of multiple activities of a complex plan. Multitudes of events compete for their attention. Commanders who do not allocate many cognitive resources to adaptive thinking will still likely feel themselves very busy; commanders who do allocate resources to adaptive thinking will need to find ways to free those resources.

Knowledge of the domain area is clearly an important requisite for performing the task well, but it is not sufficient. Typically U.S Army officers after years of study, both in the classroom and on their own, develop a good conceptual understanding of the elements of tactical decision-making. However, that knowledge alone, no matter how extensive, does not guarantee good adaptive thinking. Thinking is an active process; it is a behavior one does with his or her knowledge. As an example, if officers are told that the enemy has performed various actions on the battlefield and they are asked to infer the enemy's intent, they can generally do this fairly well depending on their understanding of the tactical domain. They have both the knowledge and the reasoning ability to solve the problem. Despite that, the same officers when placed in a demanding environment and required to perform as commanders will not necessarily display the behavior, i.e., develop a model of a thinking enemy and update that model based on continuing assessment of enemy actions. Expert adaptive thinking under stressful performance conditions requires considerable training and extensive practice in realistic tactical situations until thinking processes becomes largely automatic.

Habits and Automaticity

Habits develop only through performance. The more you repeat a behavior the more habitual it becomes, whether you want it to or not. This is true of sensorimotor behaviors such as driving a car as well as cognitive behaviors. When one is first learning how to read, for example, one's attention is focused on the shape of the letters and sounding out the words. When the sounding of the words has become automatic, one must focus attention to understand the meaning of the passage. With practice that too – extracting meaning – will become automatic. Then, one can read, understand the story, and think about what the writer is saying and whether one agrees with it. Most U.S. Army officers do not rise to a level of automaticity in battlefield thinking that permits high-quality adaptive thinking. It takes all their conscious attention to operate on the battlefield and to grasp the nature of what is happening; there are few resources left to think adaptively about the events.

Strong habits are such a critical component of expertise, in fact, that after one has attained some expertise, consciously thinking about habitual elements will usually degrade skilled performance. You cannot consciously control either thought or action with the same level of skill and complexity that you can learn to do them through repetition. Furthermore, stress narrows focus. Habits predominate in times of stress, fatigue, and competing demands for attention. Under such conditions people do what they have done most often; they do what comes automatically.

Training Methods

The cornerstone of developing expertise is the use of deliberate practice. A main tenet of the deliberate practice framework is that expert performance reflects extended periods of intense training and preparation (Ericsson, Krampe, & Tesch-Roemer, 1993). Describing the structure of deliberate practice activities, Ericsson et al. write "…subjects ideally should be given explicit instructions about the best method and be supervised by a teacher to allow individualized diagnosis of errors, informative feedback, and remedial training…. Deliberate practice is a highly structured aim; the specific goal of which is to improve performance. Specific tasks are invented to overcome weaknesses, and performance is carefully monitored to provide cues for ways to improve it further." (p. 367-8)

Traditionally the training of tactical thinking in the U.S. Army has not employed deliberate practice concepts. Instead, officers have been placed in realistic situations, supported by some form of live, constructive, or virtual simulation, and asked to perform in a whole-task environment to the best of their ability. The maxim "train as you fight" has risen to such a level of familiarity in the U.S. Army that the value of the notion goes almost unquestioned. Yet studies of the development of expertise clearly indicate that "as you fight" meaning performing in fully realistic simulated battles is neither the most effective nor efficient method of developing expertise. Such "performances" can help a novice become acquainted with applying military knowledge and can reinforce existing knowledge in an experienced person, but will not in and of themselves lead to the development of expertise. In many fields where expertise has been systematically studied, including chess, music and sports, development beyond intermediate level requires large amounts of deliberate practice (Ericsson, et al., 1993) and good coaching (Ericsson, 1996; Charness, Krampe & Mayr, 1996).

How does deliberate practice differ from exercise based on full-scale realistic performance? Here are some key characteristics that distinguish deliberate practice:

1. Repetition. Task performance occurs repetitively rather than at its naturally occurring frequency. A goal of deliberate practice is to develop habits that operate expertly and automatically. If appropriate situations occur relatively infrequently or are widely spaced apart while performing "as you fight" they will not readily become habitual.

2. Focused feedback. Task performance is evaluated by the coach or learner during performance. There is a focus on elements of form, critical parts of how one does the task. During a performance these elements appear in a more holistic fashion.

3. Immediacy of performance. After corrective feedback on task performance there is an immediate repetition so that the task can be performed more in accordance with expert norms. When there is feedback during a "train as you fight" performance, it is often presented during an after-action review (AAR) and there is usually not an opportunity to perform in accordance with the feedback for some time.

4. Stop and start. Because of the repetition and feedback, deliberate practice is typically seen as a series of short performances rather than a continuous flow.

5. Emphasis on difficult aspects. Deliberate practice will focus on more difficult aspects. For example, when flying an airplane normally only a small percentage of one's flight time is consumed by takeoffs and landings. In deliberate practice simulators, a large portion of the time will be involved in landings and takeoffs and relatively little in steady level flight. Similarly, rarely occurring emergencies can be exercised very frequently in deliberate practice.

6. Focus on areas of weakness. Deliberate practice can be tailored to the individual and focused on areas of weakness. During "train as you fight" performances the individual will avoid situations in which he knows he is weak, and rightly so as there is a desire to do one's best.

7. Conscious focus. Expert behavior is characterized by many behaviors being performed simultaneously with little conscious effort. Such automatic elements have been built from past performances and constitute skilled behavior. In fact, normally, when the expert consciously attends to the elements, performance is degraded. In deliberate practice the learner may consciously attend to isolated elements because improving performance at the task is more important than performing one's best. After a number of repetitions attending to the element to assure that it is performed as desired, the learner resumes performing without consciously attending to the element.

8. Work vs. play. Characteristically, deliberate practice feels more like work, is more effortful than casual performance, and is often less engaging and fun than fully realistic performances. The motivation to engage in deliberate practice generally comes from a sense that one is improving in skill.

9. Active coaching. Typically a coach must be very active during deliberate practice, monitoring performance, assessing adequacy, and controlling the structure of training. Typically in "train as you fight" performances there are no coaches. Instead there are observer/controllers who attempt to interfere as little as possible in the performance.

'Think Like a Commander' Themes

Repetitive performance causes behavior to become automatic. It is important that the behaviors that become ingrained conform to those of an expert—that they are the right behaviors. It is a well-known phenomenon that novices, through play alone, will improve rapidly for a short time but then may continue performing for decades without further improvement. Practice alone does not make perfect; it must be structured to ensure that performance, in this case thinking, is done in a correct manner. In order to accomplish training using a deliberate practice method the student must perform selected task elements and strive to conform his or her performance to some model of 'correct form' or 'expert form.' If those desired elements of form have not been clearly identified, then the training will resemble the discovery learning of "train as you fight" more than it does deliberate practice. A critical component in the construction of the Think Like a Commander training for tactical adaptive thinking - an explicit set of expert tactical thinking behaviors - was formulated based on ARI interviews and research with acknowledged tactical experts (Deckert, Entin, Entin, MacMillan, & Serfaty, 1994; Lussier, 1998; Ross & Lussier, 2000). These eight behaviors are termed 'themes' of the training. Below is a list of the themes and a brief description of each:

Keep a Focus on the Mission and Higher's Intent -- Commanders must never lose sight of the purpose and results they are directed to achieve -- even when unusual and critical events may draw them in a different direction.

Model a Thinking Enemy -- Commanders must not forget that the adversaries are reasoning human beings intent on defeating them. It's tempting to simplify the battlefield by treating the enemy as static or simply reactive.

Consider Effects of Terrain -- Commanders must not lose sight of the operational effects of the terrain on which they must fight. Every combination of terrain and weather has a significant effect on what can and should be done to accomplish the mission.

Use All Assets Available -- Commanders must not lose sight of the synergistic effects of fighting their command as a combined arms team. They consider not only assets under their command, but also those which higher headquarters might bring to bear to assist them.

Consider Timing -- Commanders must not lose sight of the time they have available to get things done. Experts have a good sense of how much time it takes to accomplish various battlefield tasks. The proper use of that sense is a vital combat multiplier.

See the Big Picture -- Commanders must remain aware of what is happening around them, how it might affect their operations, and how they can affect others' operations. A narrow focus on your own fight can get you or your higher headquarters blind-sided.

Visualize the Battlefield -- Commanders must be able to visualize a fluid and dynamic battlefield with some accuracy and use the visualization to their advantage. A commander who develops this difficult skill can reason proactively like no other. "Seeing the battlefield" allows the commander to anticipate and adapt quickly to changing situations.

Consider Contingencies and Remain Flexible -- Commanders must never lose sight of the old maxim that "no plan survives the first shot." Flexible plans and well thought out contingencies result in rapid, effective responses under fire.

We believe the above set of eight tactical thinking behaviors is a good set for the following reasons. First, the behaviors are characteristic of high-level expert tactical decision-makers. Observing acknowledged experts, these elements can clearly be seen guiding their actions. Second, the concepts are familiar to most officers. They have been taught to do these things and generally are able to do them with some degree of proficiency. Third, observations of officers in realistic tactical performances indicate that they typically do not perform according to these norms; the more intense the exercise, the less likely are the officers to exhibit these behaviors. Fourth, the set describes thinking actions that can be loosely characterized as "what to think about" rather than "what to think." Fifth, and very importantly, the themes represent thinking behaviors that are relatively consistent over a wide range of tactical situations. Because of that consistency, the formation of automatic thought habits will occur more quickly at this level of generality than it will for the unique and specific aspects of each situation, that is, the inconsistencies of tactical thinking.

It is not sufficient to simply memorize the eight tactical thinking themes and learn the questions that commanders must ask. In fact, as has been indicated, the eight themes are already well known in one form or another to officers at the tactical level. The themes are not intended to be a checklist either. Difficulty with adaptive thinking is a performance problem, not a knowledge problem, and it will not be solved by the acquisition of additional declarative knowledge.

'Think Like a Commander' Vignettes

This section briefly describes the design of the Think Like a Commander training product; a more extensive description is available in *Think like a commander prototype: Instructor's guide to adaptive thinking* by Lussier, Shadrick, & Prevou, (2003). Shadrick & Lussier (2002) contains the training materials used with U.S. Army captains. The central component of Think Like a Commander is a set of vignettes based on tactical situations drawn from a single overarching scenario. Each vignette begins with a short—typically two to four minutes in duration— audio-video file that presents the tactical situation.

While each vignette has no officially sanctioned solution, each does have a set of unique "indicators" that represent important considerations of expert battlefield commanders. These are the elements of the situation—the key features—that should play a role in the decision maker's thinking. For each vignette, 10 to 16 such indicators were determined. While the themes are consistent across all vignettes, each vignette has unique indicators that represent what an expert commander should consider in that specific vignette situation if he or she were to engage in the thinking behavior represented by the theme.

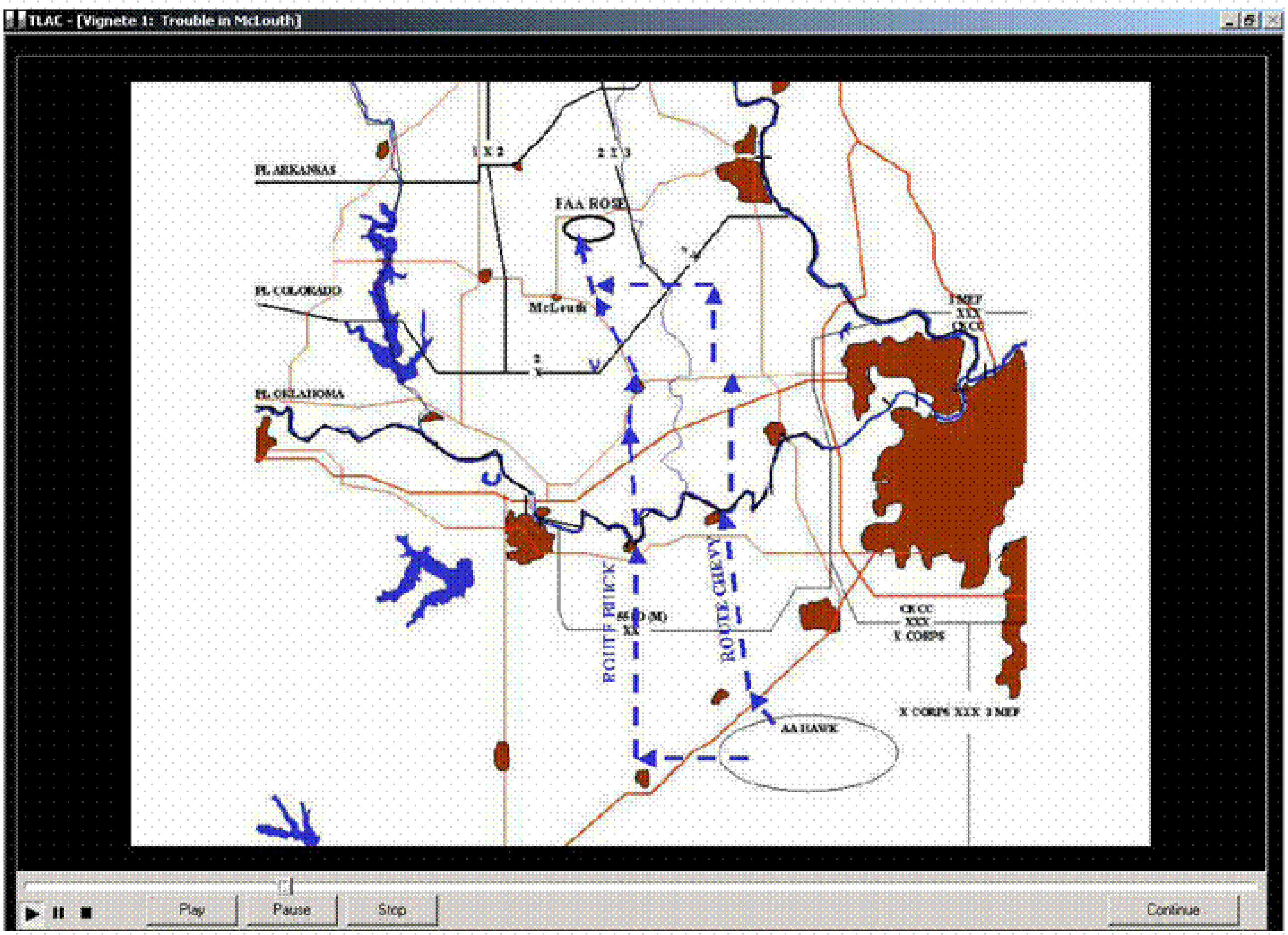


Figure 1. Think Like A Commander – Vignette Screen.

Once the presentation is completed, the student is asked to think about the situation presented and to list items that should be considered before making a decision. Over time, the instructor decreases the amount of time students are allowed, forcing them to adapt to increased time pressure. After each student completes his or her list, the instructor leads a class discussion. Class members discuss the second- and third-order effects related to actions students suggest. Students are encouraged or required to discuss and/or defend considerations relevant to the vignette. Such coaching by a subject-matter expert is a key part of the learning process to enable the student to develop expert habits.

In the final phase of each vignette, the students see the list of considerations that experts believed were important, along with the list they initially made, and they mark the indicators they have in common with the experts. Students are also asked to make the same evaluation on the class as a whole. The purpose in this step is to allow the student to get a true representation of their individual performance. For example, a student may only get fifty percent of the important considerations for a given vignette. During the class discussion, however, ninety to one hundred percent of the key considerations may be discussed. Students may inappropriately believe that their performance was directly linked to the performance of the class as a whole. Once the students rate their performances, they are given feedback linked to the general themes, (e.g., 25% for the 'Model a Thinking Enemy' theme). This individual feedback supplements and complements the feedback given by the instructor during the class discussion phase of the training. The students are then able to focus their future thinking on subsequent vignettes and place additional attention on themes for which they scored low.

Implementation and Evaluation

U.S. Army Captains in the Armor Captains Career Course at Fort Knox, KY received the adaptive thinking training using seven of the Think Like a Commander vignettes. Participants included 24 Officers enrolled in the course between January and May 2002. The training was facilitated by their classroom instructors. Each instructor received a 6-hour block of instruction on using the Think Like a Commander training. A senior instructor at the Command and General Staff College and the training program developer provided the instruction. The instruction included and involved discussion on adaptive thinking, information on how to use the materials, and techniques for facilitating an adaptive thinking discussion.

Implementation procedures were similar to those discussed earlier. After reviewing the vignette, students were asked to list all the important considerations that should be noticed from the vignette and were given a time limit in which to make their lists. Next, the instructor led a discussion of the vignette to further highlight the relevant teaching points from the vignette. Finally, students were required to complete the self-assessment section of the program. The procedures were similar for all seven vignettes. All student input was automatically saved for analysis.

Results

A number of performance measures were used to evaluate the success of the training. These measures addressed the critical thinking process (e.g., the number of critical indicators identified) and the ability to make rapid decisions (e.g., the amount of time spend determining indicators).

One key measure is the percent of critical information the student was able to identify within the time limit. The set of critical information is those items that were termed indicators in the previous section. They are a list of 10-16 items that expert tactical decision makers consider to be important considerations in the situation. A trend analysis was computed to determine trend effects through repeated use of the training application. A within-subjects trend analysis yielded a significant linear trend, $F(1, 23) = 34.21$, $p < .05$, indicating that participants identified significantly more critical information as they continued training.

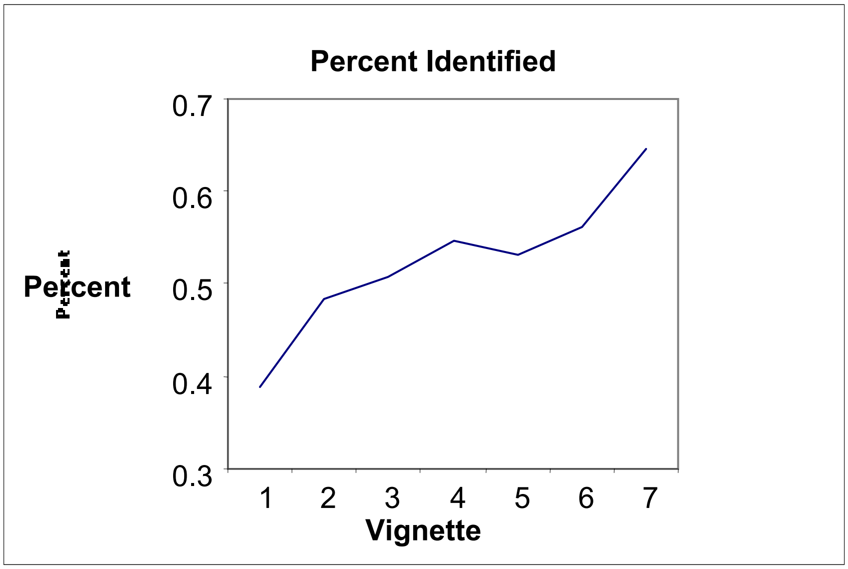


Figure 2. Percentage of Indicators Identified for Each Vignette

Students could not spend as much time as they wanted to complete each vignette. Time limits were imposed by the instructors. For the first vignette a time limit of 15 minutes was allowed for the student response portion of the exercise. The time limit was progressively reduced so that for the seventh vignette only 3 minutes was allowed. Figure 3 illustrates the amount of information considered per minute for each vignette.

For example, for vignette one, participants were allowed 15 minutes to complete the exercise and they identified an average of 6 considerations for the whole exercise, or a total of .41 considerations per minute. For vignette seven, participants were allotted 3 minutes to complete the exercise and participants

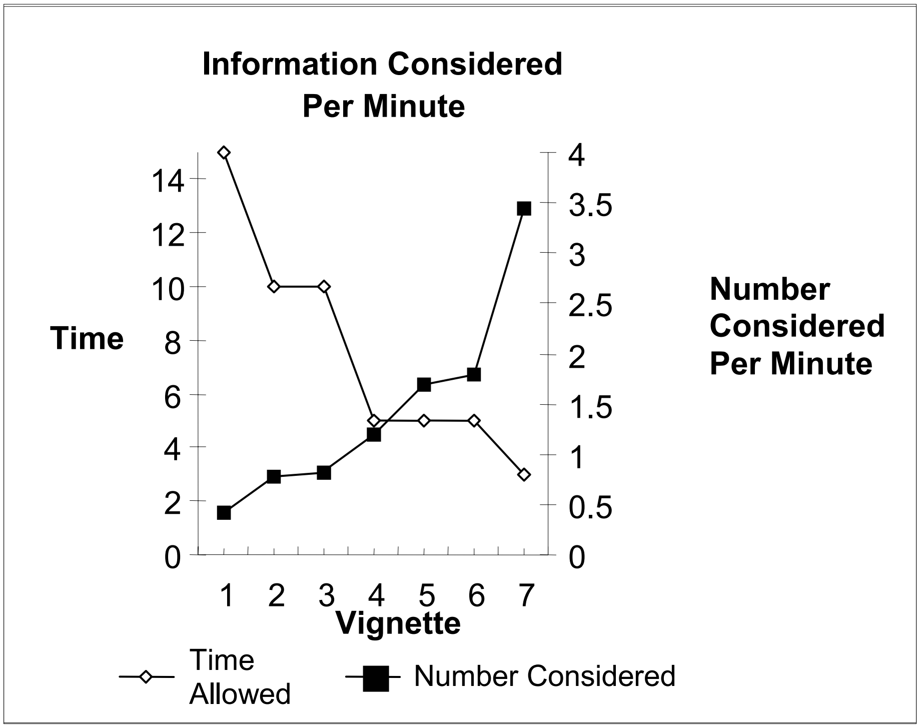


Figure 3. Percent of Information Considered per Minute.

considered just over 10 pieces of information. This corresponds to 3.4 considerations per minute. Finally, it is important to note that fatigue was not a problem for the training since students were given ample time between vignettes. Furthermore, students were motivated by their interest in the training and were actively involved in the training.

From a qualitative standpoint, both participants and instructors indicted that they enjoyed completing the Think Like a Commander training. Not surprisingly, the vignettes were able to generate lively discussion and kept the interest of the participants. Comments indicated that the students perceived the training as being valuable and many requested the software to take with them after completing the course.

Interpretation of the value of results such as these is always subjective. We feel in this case the increase is impressive particularly because each vignette presents a new tactical situation and the target considerations – the critical features that the students must identify – are also different each time. The results indicate a substantial rise in the students' ability to rapidly perceive and identify the critical features of tactical situations.

"How to Think" versus "What to Think"

Occasionally one hears a distinction made between training "how to think" and training "what to think." Indeed, previously in this paper we characterized the themes as "what to think about" as opposed to "what to think." It may be that such distinctions represent distinctions between procedural and declarative knowledge, or it may be that such distinctions, while sounding meaningful, are entirely vacuous. Recall that we do not believe we are primarily imparting knowledge with our training; rather we are shaping some fairly discrete and specific thinking behaviors.

When we were initially asked to develop adaptive thinking training it was clear that the requesters had in mind a much more portable thing, a thing in which the "how to think" and "what to think" distinction was meaningful. They envisioned a training program that turned the students into "adaptive thinkers" who, having once had the training, now thought differently, and who, as they went about various endeavors including tactical decision-making, would apply a new quality of thought to them and perform them more adaptively, whatever that might mean. While there are many adherents of generalized domain-free training in thinking skills, results of such training efforts are far from clear. In this research effort we thought it more promising to adopt a definition of adaptive thinking that was probably narrower than is often intended by the term, a definition that was very specific for the domain of tactical decision-making. We (Ross & Lussier, 1999) made the assumption that the ability to think adaptively is something that grows out of quality experiences within a domain and does not necessarily transfer readily to other domains.

Perhaps a meaningful way to approach the issue is in terms of the generality of the behavior being trained. For example, a very general thinking skill, which is applicable to a wide variety of situations, could be phrased "take a different perspective." A more specific instance, tailored to battlefield situations would be embodied in a rule such as "if the enemy does something you didn't expect, ask yourself what purpose he hopes to achieve by the act," a behavior that inclines one to take the enemy perspective. A yet more specific instance would be "when you see an enemy-emplaced obstacle, ask yourself why he put it in that exact location and what he intends to achieve by it." Recall that these thought acts – these cognitive behaviors – are not part of a large checklist that one continually seeks to proactively apply to the environment, rather they are thought habits that operate within complex structures (i.e., mental models) and must be triggered by some stimulus event. When the triggering event is very specific and identifiable such as an enemy-emplaced obstacle the training may proceed readily but has a limited applicability. Achieving the desired effect of improving adaptive thinking in tactical situations would require an enormous number of such habits be trained. At the other end of the spectrum, the mandate to "take a different perspective" is so vaguely triggered and the act of taking the different perspective so broadly flexible that a tremendous and thoroughgoing training effort must be required to achieve any lasting effect, especially when one considers the attention-demanding and focus-narrowing environment in which we seek to affect behavior. Thus, we believe the course taken in this effort to be the most efficacious one; to place the themes at just such a level of generality that they represent thinking behaviors that are as specific as possible while remaining relatively consistent over a wide range of tactical situations. Because of that consistency, the formation of automatic thought habits will occur more quickly, and because of the specificity they will more likely operate in battlefield conditions.

Refining the Themes

While the themes – which represent our model of expert tactical thinking – and our formulation of them as eight roughly equal-valued items have proven useful in practice, a recent effort which involved observation of one-on-one vignette-based tutoring sessions with an expert mentor (Ross, Battaglia, Phillips, Domeshek, & Lussier, in preparation) is continuing to refine our understanding of them. The analysis of themes in the sessions and the post-session interviews with the expert tacticians helped us to see a more sophisticated use of themes than we had first suspected. We now believe that visualization is the key skill in tactical thinking, and the other themes support it. We define visualization as the ability to integrate the use of the other themes in response to a particular situation through the activation of mental simulation. We have observed that the basic themes, the adaptation of the themes to understand new situations, and mental simulation, which integrates the themes into a holistic view of the battlefield, can all be tutored and that practice using the themes in context creates deeper understanding.

Ross et al. have also begun to further develop the evaluation of the theme-based behaviors by developing behaviorally anchored rating scales, as for example in Figure 4.

T1: Keep a focus on the Mission and Higher Intent.
Commanders must never lose sight of the purpose and results they are directed to achieve – even when unusual and critical events may draw them in a different direction.

Focus on Own Mission	Discriminate Intent and Explicit Mission	Model Effect of Own mission on HQ Intent	Accurate Predictions	Support Intent
1	2	3	4	5
Articulates and understanding of the mission without any consideration of higher intent Neglects to keep HQ informed of plans and situations Neglects to request additional assets when the plan requires Ignores of loses sight of higher intent when distracted by unusual events	Is able to differentiate mission from higher intent, yet does not apply these differences to understanding the current situation in front of him Understands both mission and intent, but does not consider whether mission will support that intent or whether it needs to be modified in any way to better support intent	Considers whether the mission will support the intent Considers whether mission needs to be modified in order to better support the intent Considers ways to modify mission to better support intent Thinks through what has to be accomplished in order for the higher intent to be achieved	Mentally simulates how his mission will contribute to achieving a larger mission Prioritizes what needs to happen in order for the higher mission to be accomplished (e.g., I need to do this, instead of that.)	Articulates how and/or why his plan or COA supports the commander's intent Allows intent and current situation to guide the COA rather than the explicit mission

Figure 4. Behaviorally Anchored Rating Scale for Keep a Focus on the Mission and Higher's Intent Theme.

Future Directions

Two important and related considerations are transfer of the skills trained by Think Like a Commander to more realistic whole-task situations and degree of automaticity of the thinking behaviors. Neither an assessment of transfer nor of automaticity has been accomplished. It is unlikely that seven short vignette-based exercises involved sufficient repetition to reach even a low degree of automaticity. It has been noted, however, that instructors who participated in the training evaluation continued to provide coaching based on the themes as students participated in other exercises, including more complete simulation-based exercises. Both automaticity and transfer are a result of continued production of the behaviors performed in the training vignettes in a variety of tactical exercise settings.

In additional research, we are investigating the pattern of development of tactical thinking related to the themes. If the development of tactical thinking skills follows a consistent and discernable pattern then individual performance levels can be diagnosed, and training can be more efficiently targeted to individual needs. One consistent finding from a number of efforts (Deckert, et al., 1994; Ross, et al., 2003; Carnahan, Lickteig, Sanders, & Durlach, in preparation) shows that in novices the amount of attention focused on own forces, i.e., the theme Use All Assets Available, is much higher than the amount of attention placed on the enemy, i.e., the theme Model a Thinking Enemy, but becomes more balanced or reverses with the development of greater expertise. The finding has been noted in other fields, e.g., chess, where it is a frequent observation that novices focus on their own plans and moves and seem to ignore what the opponent is doing. One explanation is that in order to act one must consider 'own forces,' and such consideration virtually exhausts the capacity of the novices to build, maintain, and operate their mental models. Only with increasing expertise are models of sufficient complexity to encompass both 'own forces' and 'enemy forces' possible. Another explanation (Ross, et al.) is based on the tendency of novices in all domains of expertise to jump to solutions before gaining a sufficiently deep understanding of the situation.

Other research based on Think Like a Commander training development focuses on the method of delivering coaching. Good coaching is seen as an integral part of the training method. In the work reported here instructors performed the coaching, i.e., live coaches were employed. Other research efforts currently underway are investigating various alternative methods of delivering the coaching component, including intelligent tutors, non-interactive presentations, live but distant coaches, asynchronous interaction with a live coach, and collaborative student groups.

References

Carnahan, T.J., Lickteig, C.W., Sanders, W.R., & Durlach, P. J. (in preparation). *Novice versus expert command groups: Preliminary findings and training implications for future combat systems* (ARI Research Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Chaness, N., Krampe, R., & Mayr, U. (1996). The role of practice and coaching in Entrepreneurial Skill Domains: An International Comparison of Life-Span Chess Skill Acquisition. In K. A. Ericsson (Ed.), *The Road to Excellence.* (pp. 51-80). Mahwah, NJ: Erlbaum Associates.

Deckert, J. C., Entin, E. B., Entin, E. E., MacMillan, J., & Serfaty, D. (1994). *Military decision-making expertise: Final report*. Fort Leavenworth, KS: U. S. Army Research Institute.

Dvoretsky, M. (1996). *Positional Play*. London: Batsford, B.T., Ltd.

Dvoretsky, M. (1997). *Secrets of Chess Training*. London: Batsford, B.T., Ltd.

Ericsson, K. A. (1996). The Acquisition of Expert Performance. In K. A. Ericsson (Ed.), *The Road to Excellence.* (pp. 1-50). Mahwah, NJ: Erlbaum Associates.

Ericsson, K. A., Krampe, R. Th., & Tesch-Roemer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363-406.

Lussier, J. W., Ross, K. G., & Mayes, B., (2000). Coaching techniques for adaptive thinking. *Proceedings from the 2000 Interservice/Industry Training, Simulation, and Education Conference.*

Lussier, J. W. (1998). *Developing expertise in battlefield thinking: Models of battlefield thinking* (Unpublished briefing). Ft. Leavenworth, KS: U. S. Army Research Institute.

Lussier, J. W., Shadrick, S. B., & Prevou, M. I. (2003). *Think Like a Commander prototype: Instructor's guide to adaptive thinking* (ARI Research Product 2003-01). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Ross, K.G., Battaglia, D., Phillips, J., Domeshek, E., & Lussier, J.W. (in preparation). Mental Models Underlying Tactical Thinking Skills. *Proceedings from 2003 Interservice/Industry Training, Simulation, and Education Conference.*

Ross, K.G. & Lussier, J.W. (1999). A training solution for adaptive battlefield performance. *Proceedings from 1999 Interservice/Industry Training, Simulation, and Education Conference.* [On-line]. http://www.iitsec.org.

Shadrick, S. B., & Lussier, J. W. (2002). *Think Like a Commander: Captain's Edition – Prototype 1.0*. (ARI Research Product 2002-02) Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Soltis, A. (1994). *The Inner Game of Chess*. New York: David McKay Company, Inc.

Using Integrated Synthetic Environments To Train The Command And Control Of Military Assistance To Civil Emergencies

Mr M. Kelly
QinetiQ, Cody Technology Park,
Farnborough, GU14 0LX
UK
mkelly1@QinetiQ.com

Dr N. Smith
QinetiQ, Cody Technology Park,
Farnborough, GU14 0LX,
UK
nasmith@QinetiQ.com

www.QinetiQ.com

Summary

A research programme has been initiated to develop an integrated SE for training the Command and Control of civil emergencies. The focus for this work is on the UK Military Assistance to the Civil Authority (MACA) activities. Events such as the fuel crisis, strikes by fire and ambulance services, and the increasing demands generated following the World Trade Centre attack, have all required the deployment of MoD assets. The SE research community in the UK has long wished to see its applications migrate from military use to the civil domain. Therefore this study seeks to identify where existing SE tools and techniques can be used to support military aid to civil authorities.

This paper discusses the research programme, sponsored by the UK MoD Corporate Research Programme (SSS Domain) and describes the work carried out to date. Research conducted so far includes a stakeholder analysis of recent national crises. This analysis led to a generic model of crisis development, which identifies five phases ranging from prevention through containment to a review phase at the end. The application of SEs to support these phases is then described. In parallel to the stakeholder analysis, a survey of simulation tools that have the potential to support crisis management was performed and is described here. Finally some prototype architectures are illustrated and an overview of future work in this programme is discussed.

1. Introduction

The UK MoD Synthetic Environment Co-ordination Office (SECO) have long had the desire to see the research in synthetic environments migrate from the applied military research domain to other key areas. The rise in concern over terrorist attack since 9/11 and civil emergency situations in recent years increased the desire for a research programme to support the MOD department of state function. The current programme "Using Synthetic Environments (SEs) for Department of State Activities", MOD reference 20483, was therefore initiated. The focus of this study is to provide an integrated SE to support scenarios where the armed services are required to contain civil emergencies. In the UK military support to civil emergencies falls under the term Military Aid to the Civil Authorities (MACA). MACA is a catchall term that includes the following acronyms:

- MACP - Military Assistance to the Civil Power (e.g. troops deployed in Northern Ireland);
- MACC - Military Assistance to the Civil Community (e.g. flooding);
- MAGD - Military Assistance to Government Departments (e.g. petrol crisis).

This is a two year programme which in the first year seeks to understand the MACA domain, identify relevant stakeholders and assess the applicability of SEs to support MACA operations. In the second year the programme is focused on developing a prototype architecture to integrate SE modelling tools which can support MACA.

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

What do we mean by SE?

The study seeks to utilise the wide range of Synthetic Environment knowledge and technology that has been produced in the last decade of research and development to support military applications. A number of definitions have been put forward to describe what an SE is, including the SECO definition[1] - *"A computer-based representation of the real world, usually a current or future battle space, within which any combination of 'players' may interact. The players may be computer models, simulations, people or instrumented real equipments"*. Therefore, throughout the course of this paper the term SE takes a very broad description. Modelling data in the live, virtual and constructive domains are included in this definition as well as electronic conversions of data held in more traditional spreadsheet and tabular formats.

2. Stakeholder analysis

Conducting an analysis of published inquiries[2,3,4,5] and interviews with stakeholders enabled us to understand the range and scope of the agencies and organisations involved in civil emergencies. The diversity of stakeholders and the related information that they require, or may generate, reinforces the potential complexity of the data a SE based crisis management tool needs to integrate. A number of case studies taken from recent UK crises that fall into the MACA category have been studied. The case studies enabled the identification of stakeholders, information needs and potential crisis monitoring metrics. The three case studies chosen were the foot and mouth outbreak 2001, the petrol crisis 2000 and the fire fighters strike 2002/3. A brief overview of each crisis follows.

Case Study 1. Foot & Mouth 2001

The first case of foot and mouth was confirmed on 20th February 2001. On the 9th March the lead government agency the Ministry of Agriculture Fisheries and Food (MAFF) formerly requested assistance from the Royal Army Veterinary Corps and other MOD assets to assist in the planning and logistics aspects of combating the disease. The disease was finally eradicated by the end of September 2001. Over 100 agencies and pressure groups were involved throughout the course of the crisis.

Epidemiological models were used during the crisis to predict the spread of the disease and in turn assist the development of the management strategy. This strategy transitioned into an aggressive culling policy that resulted in the slaughter of over 4 million animals for disease control purposes. The number of confirmed cases per day provided a clear metric of crisis evolution.

Case Study 2. Petrol Crisis 2000

The petrol crisis occurred for one week in September 2000 and briefly again in November. Driven by the increasing costs of petrol in the UK the newly formed Farmers For Action (FFA) pressure group led a series of blockades targeted at oil refineries. The petrol shortage at filling stations was exacerbated by panic buying from the public. Initial public support faded once the health service began suffering and occupational health workers were affected. The military were on hand to drive tanker trucks if required however. The FFA called off the blockades with a 60 day warning of repeat action if no concessions on fuel prices were made. Due to minor concessions by the Chancellor and memories of the last few days of the September crisis, the repeat action did not receive sufficient support and the crisis failed to materialise.

Although no military involvement was required SEs could have been used to support the planning and management of the petrol crisis. Models of the critical communications infrastructure (road and rail) with an overlay showing refinery locations and fuel stocks would provide a useful visualisation of the crisis unfolding.

Clear performance metrics for this crisis were fuel station stocks and potentially peak time queues at filling stations.

Case Study 3. Fire Fighters Strike 2002/3

When pay negotiations between the Fire Brigades Union (FBU) and local authorities broke down, a series of strikes by fire fighters was initiated. The Office of the Deputy Prime Minister requested assistance from the MOD to provide emergency cover to contain the crisis. The first strike was scheduled for 29th October giving the armed forces approximately two months to prepare and train the fire fighter crews to use specialist equipment such as the green goddess pump (a 4 wheel drive fire engine designed in the 1950s). Joint command centres were established in local police force headquarters and staffed by senior police, fire and armed forces personnel. The role of the command centre was to filter the emergency calls and dispatch appropriate resources. Typically this took the form of initial on-site inspection by police vehicles followed by military fire fighting vehicles if the incident posed a threat to human life. The use of these joint command centres was seen as a success and indicates that a combined command centre (potentially augmented by SE visualisation) would be beneficial in general crisis management. The first strike actually occurred on November 13th for 48 hours with further strikes in November, January and February.

The fire fighters strike produced two key training requirements on military personnel. The first was just in time training of specialist equipment and vehicle training for handling the green goddess vehicle. The second was joint command team training between military and civil officers. Using SEs to support command team training in the military domain is well established. Modifying these training systems to reflect the joint resource management aspects of this strike is achievable. The time to respond to fire incidents would provide a useful metric of the effectiveness of the emergency cover and could be captured easily in both real world and simulated domains.

The studies into the three historical crises in which MACA was applied or was ready to be applied, led to two outcomes: a model of crisis development, and a rich picture of a generic MACA crisis[6].

3. Model of Crisis Development

A model of crisis development can be used to quantify the varying role of agencies compared to stages of crisis and enable emergency planners to switch resources proactively. The model developed and shown in figure 1 identifies five key stages;

- **Prevention** - The objective is to stop a crisis occurring in the first place. This objective is achieved by gathering intelligence about emerging threats and taking preventative measures as appropriate.
- **Preparation** - Applying resources to stop a known crisis causing damage. This phase includes planning, training, stockpiling vaccines etc.
- **Containment** - remedial action to minimise the effects of a crisis event or damage limitation
- **Recovery** - restore conditions to normal after a crisis event and re-building after the damage caused by a event
- **Review**.- Lessons learned (does not always happen) which can be used to improve training

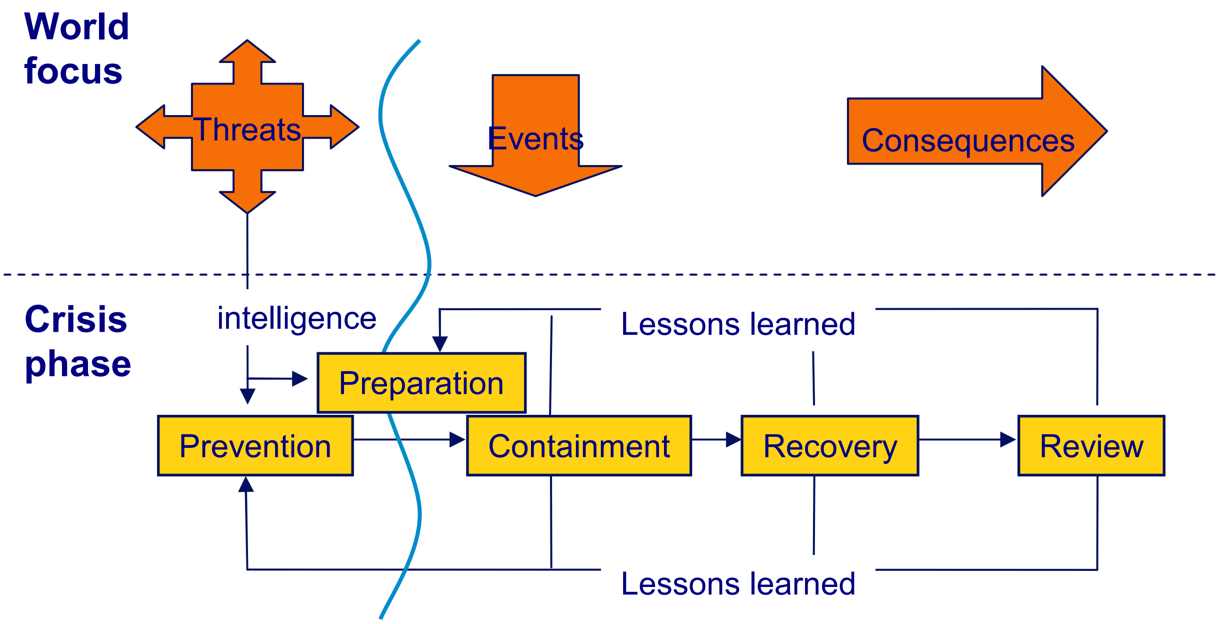


Figure 1. Model of crisis development

There is a degree of overlap between the prevention and preparation stages due to the uncertainty of when a threat transitions into a crisis event. One can imagine a use for SEs across all these phases from 'gaming' future crises in the prevention stage to an SE based AAR in the review stage providing enhanced visualisation of the evolution of an incident. It is important to develop a scalable solution and an open architecture in any developed system to customise the SE(s) to represent a specific crisis.

4. Crisis Metrics

To control a crisis and manage resources effectively it is helpful to understand where in the crisis model you are, at a point in time. A key to understanding what stage of a crisis you are in, is the capability to identify and monitor valid metrics. In foot and mouth outbreak the confirmed cases in the animal population was a clear metric. In other operations such as the return to normality in Kosovo following the conflict, the frequency of taxi use by the civilian population was another useful metric. Queue length at petrol stations and petrol station stocks may have been used as useful metrics during the petrol strike. Time to respond to a fire incident would have been a useful metric during the fire fighter's strike. This metric would measure how effective the emergency cover operated. To choose an appropriate metric one must consider a number of factors:

- The metric should be validated;
- More than one metric should be used where possible;
- Chosen metrics may not at first seem an obvious choice;
- Different metrics may be required for different stages of the crisis development.

Knowing the stage of a crisis is important because that awareness allows planners to switch resources proactively, thereby saving time & money. Figure 2 illustrates how a generic crisis metric may evolve over time superimposed over data from the foot and mouth outbreak. The data points reflect the number of confirmed cases per day recorded by the MAFF (now called the Department of Environment Food and Rural Affairs (DEFRA)).

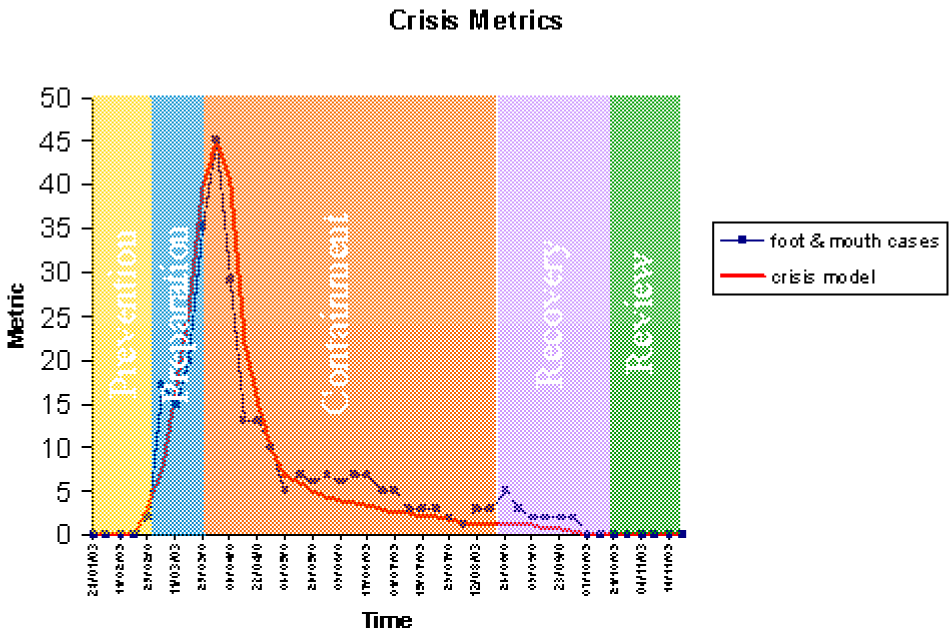


Figure 2 Generic crisis metric compared with foot and mouth cases per day

5. Application areas for SEs in Crisis Development

From the generic model of crisis development and the discussions on crisis metrics we can consider two main applications where SE's may usefully support the crisis management process.

- SE Command and Control Centre – use SEs to provide an aid to the decision making process through crisis monitoring and visualisation (applicable across all phases). This could either use live data feeds for management or be stimulated by a SE crisis simulation centre for training or operational analysis (OA).
- SE Crisis Simulation – an underpinning simulation to support both OA to test emergency measures (in the prevention phase) and training for crisis management (for the preparation phase).

SE Command and Control Centre

SE can be used to support the command and control of crises. The Command centre could be developed along military lines with external links to other key stakeholders and cells for the current operational picture, forward planning, media, police, fire army et cetera. SEs have been used to train military headquarters and unit commanders for some time. A logical progression is to extend this concept to train joint military and civil agencies (police, fire). Such a command centre would include the following key elements;

- Large screen display for shared situational awareness
- Closely coupled specialist cells
- Teaming with key resources (armed forces, police, fire)
- Scalable map display showing varying levels of resolution (e.g. vehicle to unit level aggregation).
- An underpinning architecture to integrate and convert heterogeneous data sources into an electronic representation.

The SE in the command centre would provide enhanced visualisation for command centre using 2/3D modelling and enable the visualisation of live assets. The integration of live assets into the constructive domain has been demonstrated in previous UK MOD sponsored programmes using the BATS/DIS[1] interface. The aggregation and dis-aggregation of entities has been demonstrated under the Joint Intelligence Virtual Environment (JIVE)[7]. This capability would be useful in providing a scalable picture of deployed assets from the top-level aggregate view to a more detailed entity level view.

The SE command centre could also provide an After Action Review (AAR) facility. AAR is a formalised system of providing feedback to trainees to improve their skills. The use of AAR in military training is well established. AAR for civil crises are less well understood as they require the identification of additional metrics of command staff and emergency planner performance. Using a SE based AAR with objective performance metrics would provide a clear method to visualise emergency plans, their actions and associated consequences. Finally, lessons learned from the crisis can be used to update training scenario databases.

SE Crisis Planning / Training through Simulation

In order for SEs to support the prevention and preparation phases of a crisis, an underpinning simulation of the containment phase is required. This simulation would be used to stimulate the emergency planners and command and control staff for OA or training applications. Clearly a SE is of limited use as an aid to lifting sandbags to build flood defences or controlling a fire hose. However, there may be some benefit in developing computer based training (CBT) applications for specialist equipment training.

The established SE technology of a distributed environment would allow large scale training exercises to be developed including geographically dispersed agencies such as local authorities. Emergency planning officers in local authorities may be involved infrequently in a crisis but still have a role to play (e.g. authorising the release of sandbags for flood prevention). In some situations where these individuals could not participate in such an exercise, agent based software could be developed to take the roles of the additional stakeholders. A method for incorporating these infrequent stakeholders is required that does not impact on the smooth running of the crisis simulation.

Changes are required to existing SEs to accommodate the additional requirements of MACA scenarios. Civil assets need to be represented including police and fire vehicles and their associated communication systems. This includes representing the vehicle systems and the behaviours to perform simple tasks in SAF based systems. Containment models are also required such as the impact of sandbag banks on river flow or water hoses on building fires. In short much greater fidelity is required to link the impact of people and platforms to the environment.

In all crises the ‘public’ can have a great deal of influence in the recovery from an incident. Population models have a role at the incident itself through the generation of victims (displaced population, evacuations etc.) and at the highest strategic levels (response to a remedial course of action can have political consequences).

Additional uses of a crisis simulation system include;

- A test harness for metric evaluation or highlighting potential metrics.
- Hypothesis testing in the recovery phase. This could be to analyse the re-allocation of resources in order to restore conditions to normal faster.

6. Data requirements for SE systems to support Crisis Management

[1] BATUS Asset Tracking System / DIS interface. This software takes live vehicle positions from an asset tracking system and converts the data into DIS packets, facilitating live/constructive interoperability.

Table 1 shows some data requirements drawn from the case studies and an additional case – flooding, compared to their implications for SE based monitoring of a live event and a simulated incident for training or OA.

Incident / Data Sources	Synthetic Environment LIVE (monitor)	Simulation (Training / OA)
Fire Strike		
Vehicle locations Weather Fire dispersion Chemical data (risk assessment) Road usage Road network Hospital locations	Asset tracking to DIS Link to weather server Feedback from field view hazards, handling procedures Live update from traffic associations (AA, RAC) Update from councils for road closures Map overlay for nearest location	Semi-Autonomous Force (SAF) models Simulate weather patterns (SAF systems) Model of fire dispersion Simulate chemical fires Traffic Simulation on city wide scale Refine simulation with obstacles Add features to database
Foot & Mouth		
Epidemiological model Foot & mouth cases No of VETS / military vet service?	Visualise algorithm Live update from field Display resource as table	Implement algorithm in simulation Use algorithm for dispersion simulation Computer generated vets?
Petrol Crisis		
Fuel supplies Road network Refinery locations	Update from field Rolling road block protesters Map overlay	Logistics simulation Simulate communications (road and rail) infrastructure Add features to database
Flooding		
River capacity / water table / Rainfall Tide information Demographic data (population at risk)	Tabular data converted to 2/3D maps Tables / Satellite / aerial images Represent as GIS layer	Need model with high resolution (~inch resolution in height) 3D simulation of flooding GIS based Advanced population models for evacuation planning

Table 1. Data requirements mapped to their implications for SE use.

There are a number of common key data requirements that it can be assumed would be required by any SE system that might support crisis management.

These include:

- Scalable map data from large area rural to city and town maps;
- Maps of special services such as underground, rail and air corridors;

- Mapping of utilities (gas and electricity);
- Vehicle model data;
- Time dependant data such as disease spread, traffic flows, crowd behaviours;
- Meteorological data;
- Traffic and population movement models.

These components indicate a number of discrete tools that will form part of the crisis simulation architecture.

7. Crisis simulation architecture

The review of case study material approach was used to develop the rich picture of a crisis incident (figure 3). This picture highlights the complex organisational structures and data flows at work during a crisis event. This complex picture indicates the requirements for a SE to incorporate heterogeneous data sources in any potential system for crisis management.

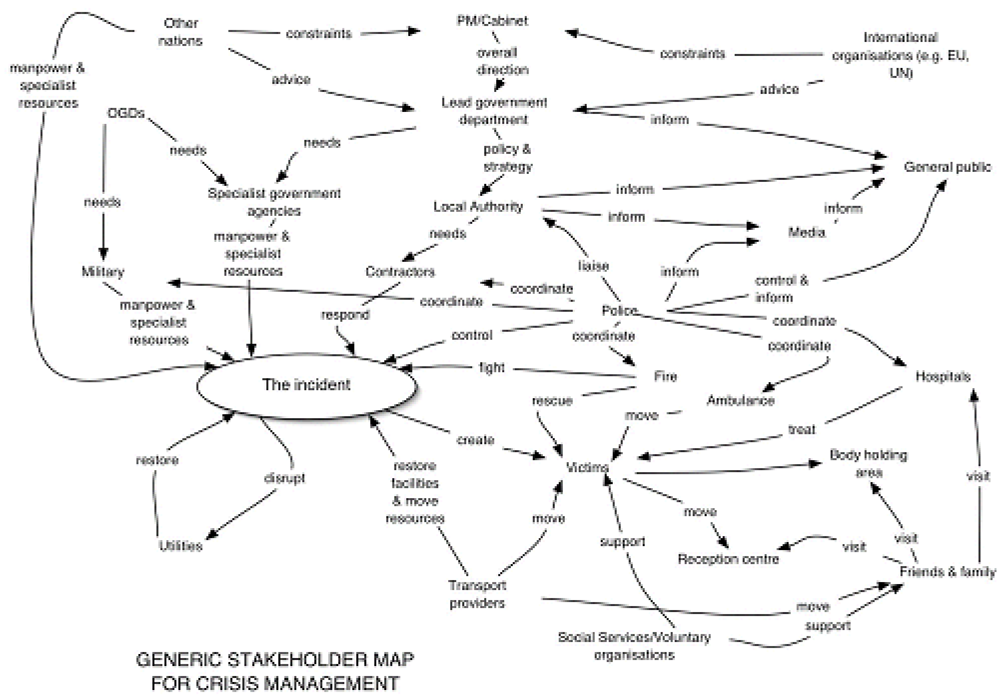


Figure 3. Rich Picture of a generic incident

Survey of Potential Tools

At the start of this research programme a number of tools were known to the authors that were applicable to crisis management. Therefore, rather than design an integrated SE for crisis management from the ground up it was decided to re-use current simulation applications where possible. A search has been undertaken to identify SE applications that can be re-used in a larger SE for crisis management. Due to the diversity of MACA scenarios the search encompassed a wide range of application areas. The search highlighted approximately 60 applications with the potential to support a MACA SE. These range from CGF applications, familiar to the military community, virtual trainers for first responders and simulations of natural phenomena such as flooding. Examples include ATLAS Ops (a police force command & control tool) Diablo VR (a first responder trainer) ANFAS (a GIS based flood analysis tool).

An issue with these applications is the ease with which they can be networked to play in the same environment and the modifications required to support MACA scenarios. Modifications would include introducing the dynamics of civil assets and the "weapons" systems i.e. how to simulate a fire hose and the dynamics of containing a fire. Some scenarios may require simulations of large numbers of civilians. Successful applications of this kind have been demonstrated at QinetiQ using fluid dynamics to represent the population and a DIS interface to allow for interoperability.

If a SE is constructed to simulate the vast range of MACA scenarios, a composable architecture is required which can be tailored for a specific incident. This "golf bag" approach is needed so tools can be selected and plugged into a system architecture to provide the underpinning simulation. Figure 4 displays a hierarchical concept where simulation tools comprise HLA federations at three levels (Tactical, Operational and Strategic). These levels represent the management levels that operate during civil crises. This architecture arises from the rich picture of a generic incident and the stakeholder analysis carried out previously. Each level is a federation in its own right. Communication between levels is facilitated by an HLA gateway. This approach would allow close coupling between the components that constitute the simulated incident. Additional players at a higher level could take part in the simulation through the gateways. For example the local authority planning officer could dial into the simulation through a secure server to participate infrequently through the course of the exercise.

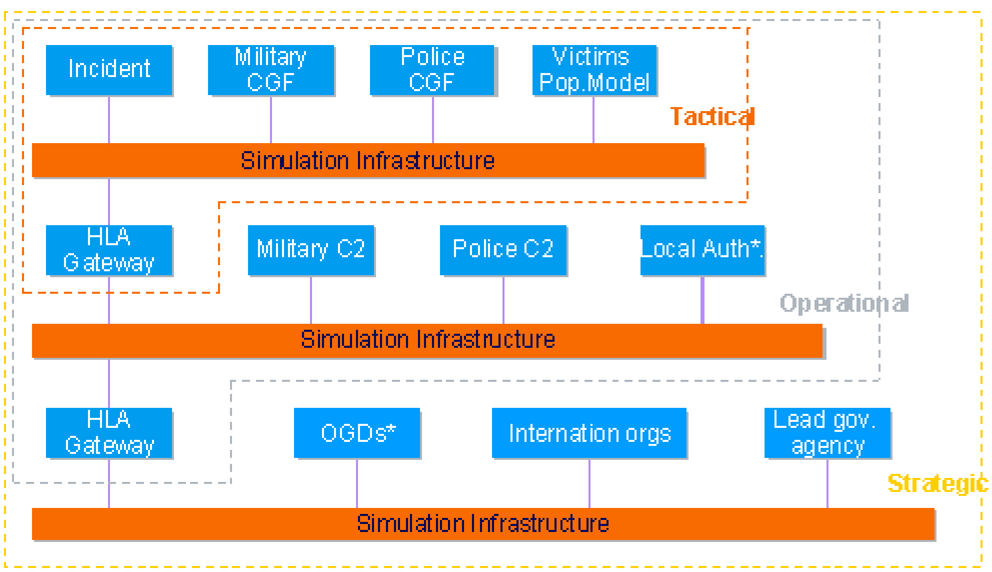


Figure 4 Hierarchical simulation architecture

* Local Auth. = local authorities, OGD = Other Government Departments.

8. Future Work

The hierarchical simulation architecture represents the first pass at a simulation concept. The next stages of this programme will develop the SE command centre and crisis simulation concepts further. A number of scenarios will then be developed to focus research in the second phase of the programme. A shortlist of tools will then be selected to represent the required components of the crisis simulation system. It is anticipated that this crisis simulation will be used to stimulate a demonstration of the SE based command and control centre to MoD and Other Government Departments.

9. Acknowledgements

The authors would like to thank the UK MoD Synthetic Environment Co-ordination Office for sponsoring this work and Gregory-Harland Ltd for supporting the stakeholder analysis study.

10. References

1. SECO definition of Synthetic Environment www.mod.uk/issues/simulation/definition.htm
2. The Anderson Inquiry, Foot and Mouth Disease: Lessons Learned Inquiry, 2001
3. The Royal Society Infectious Diseases in Livestock, 2001
4. The National Audit Office, The Outbreak of Foot and Mouth Disease, 2002
5. Hathaway P., The effects of the fuel protest on road traffic, Transport Statistics Road Traffic Division, DEFRA, 2000.
6. "Synthetic Environments in support of crisis management", Gregory Harland Limited February 2003
7. Joint Intelligence Virtual Environment (JIVE) International Training and Education Conference, Lille, 2002.

Augmenting Electronic Environments for Leadership[1]

Joseph Psotka
Ken Robinson
US Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Lynn Streeter
Thomas Landauer
Karen Lochbaum
Knowledge Analysis Technologies
Boulder, CO 80301

Summary

Adapting instructional content to match the background knowledge of the student has been a long-standing goal of student modeling and tutoring (Bruner, 1966; Burton & Brown, 1979). To this end, cognitive scientists have developed student models that rely on manual knowledge bases relevant to the particular instructional task at hand. Normally, these include domain content, instructional content, models of student misconceptions, and more. While tailoring instruction to the learner has been shown to be effective (Anderson, 2002), current approaches are difficult to implement because of the enormous amount of skilled professional effort required.

Ideally, the system should *automatically* select the most appropriate content for the student based on a minimal amount of student data. This educational desideratum has been coined the "Goldilocks Principle"---providing lessons, texts, probes, etc. that are neither too advanced nor too elementary, but just right—in the "zone of proximal development (ZPD)" (Palincsar & Brown, 1984; Vygotsky, 1978), known to be important for promoting efficient learning.

In an early study, Wolfe et al. (1998) used Latent Semantic Analysis (LSA) to automatically select the next text passage for students to read and achieved one-sigma learning augmentation effects (Bloom, 1976). In the study reported here, we examined the Goldilocks Principle in the context of an LSA - enhanced online discussion environment, where contributions were automatically selected to be most similar in meaning to learners' notes or contributions. We found that these contributions were almost always of somewhat higher judged quality. We suggest that this implementation of the Goldilocks Principle is a consequence of how LSA represents consensual knowledge, and provides an automatic way for selecting the next best piece of material for a student to learn, making this an important contribution to tutoring.

INTRODUCTION

Student Modeling and Tutoring using Latent Semantic Analysis

Automatically adapting the content of instruction to students' background knowledge has long been the holy grail of student modeling and tutoring (Bruner, 1966; Burton & Brown, 1979). Four types of cognitive models have been developed for use in interactive intelligent tutors: global descriptions in terms of components such as cognitive and learning styles; overlay or tracing models specifically describing the micro units of instruction; models of student errors and misconceptions; and models that structure the knowledge to be learned at graduated depths of expertise (Psotka, Massey, and Mutter, 1988). The Achilles' heel of all these approaches is the large cost of developing the models. As effective as they are, they are very difficult to implement, largely because of the enormous professional efforts required to organize the various knowledge bases for subject matter content, instructional strategies, student assessment of bugs and misconceptions, and knowledge levels. To move student modeling to a more practical plane, automated

[1] This paper does not represent US Army official policy. Approved for public release. Please contact the lead author, Dr. Joseph Psotka, US Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Ave., Alexandria, VA 22333 at Joseph.psotka@us.army.mil.

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

methods of locating and constructing appropriate knowledge, and to code, test, and assure the effectiveness of the representation, delivery, and interaction software are needed.

Latent Semantic Analysis (LSA), a machine learning technology for simulating human meaning of words and text passages appears to be appropriate for the student modeling and tutoring problem. LSA is both a model of human knowledge representation and a method for extracting and representing the meaning of words mathematically (for greater detail, see Landauer and Dumais, 1997; Landauer et al., 1998; Landauer, 2002). LSA induces word and passage meanings by mathematically analyzing a large corpus of relevant text. In the result, every word and every passage are represented as points in a high-dimensional "semantic space." This space defines the degree of estimated semantic similarity between any two words or passages. Simulations of many linguistic, psycholinguistic, and human learning phenomena, as well as several educational applications, show that LSA very accurately reflects corresponding similarities of meaning as judged or used by humans (Landauer and Dumais, 1997; Foltz et al., 1999). As we will show, LSA also provides an expeditious approach to the creation of tutoring models, by learning how to relate the semantic similarity of many different written descriptions and written answers to complex problems, and combining and presenting them automatically in effective sequences.

Wolfe et al. (1998) applied LSA directly to the problem of finding the next best piece of text for a student to read. They reasoned that the ability of a reader to learn from text depends on the match between the background knowledge of the reader and the difficulty of the text information. They used LSA to measure the distance between the reader's knowledge as gleaned from their short essays and graduated essays on the structure and function of the heart, collected from four sources of increasing difficulty.

In their study, college and medical students wrote short essays on the anatomy and function of the heart. The students next read one of four texts that ranged in difficulty from elementary to medical school level, and then wrote a new essay. Results showed that learning was greatest for texts that were neither too easy nor too difficult. Essays were represented in an LSA space which included articles on the heart and circulatory system. Degree of difficulty between the essay and a piece of text was indexed by the similarity between them as measured by the cosine in the same space. For each essay the vector in this 100 dimensional space was found and compared to the four target texts. A low cosine value between the text and essay would indicate low similarity, and thus reading this text would produce little learning. On the other hand, a too high cosine would indicate that the student already knew the content of text, and thus would be unlikely to learn anything new from the text passage. Learning, as measured by pre – post gains in short answer tests and essays (independently scored by ETS), was greatest for intermediate cosine similarity values. A more advanced version of the method (Rehder et al., 1998) placed all the essays on a line, so that cosines indicating similarity of texts that were less advanced and those that were more advanced could be distinguished. This produced even stronger results.

The LSA cosine similarity measure proved as effective at predicting learning from these texts as the traditional knowledge assessment measures. The implication is that, like Goldilocks, a student should be offered explanatory material that is neither too easy nor too difficult, but just right. Optimally the texts should stretch the student's understanding but still be comprehensible, and introduce some, but not too many new concepts. Typically in the Wolfe et al. results, the best learning occurred with texts whose cosine distances from student essays were in the range of 0.5 to 0.6. Although the magnitude of these cosines is particular to the dimensionality of each singular value decomposition semantic space, it is instructive that they lay somewhere in the middle of the positive range.

Applying Goldilocks to Collaborative Learning

In the study reported in this paper, the Goldilocks Principle was applied to contributions in a customized electronic discussion environment. If a broad range of novices, journeymen, and experts write about their approaches to solving a problem, it should be possible to analyze their contributions systematically and automatically, using LSA, to provide graduated responses that vary in conceptual difficulty, complexity, and number of constituent themes that compose solutions. Although in this work, we were not yet able to vary the responses dynamically during discussions, we were able to analyze the spontaneous interactions in ways that serve to evaluate the hypothesis. The analyzed interactions were discussion contributions written by US Army officers who varied in rank and range of expertise. Contributions were later separately rated for overall quality by expert raters. We then examined whether the spontaneous notes of other officers, presumably ones offered on the basis of different knowledge and intent, could be used to implement the Goldilocks Principle. The goal was to find an automated way to find the next best discussion contribution or text sample for the student to read—something that was just right. As a first order approximation, the system did dynamically point the officers to other contributions that were most similar to theirs, as measured by LSA.

Problem Scenarios

Tacit Knowledge of Military Leadership (TKML)

Four different scenarios dealing with military management situations were used in the online discussion environment. The scenarios were developed by Yale University, in collaboration with the Army Research Institute to assess Tacit Knowledge of Military Leadership (TKML) (see Hedlund et al., 2000; Sternberg et al., 2000). The method is based on a carefully developed set of representative scenarios of challenging interpersonal leadership situations that are commonly encountered by Army officers, along with sets of alternative actions that a leader might take. Interviews with experienced officers originally suggested the scenarios and alternatives. Scenarios are refined, tested, and edited through a process of expert judgments, trial uses, and evaluation against a criterion of showing more expert-like responses with increasing military rank. In previous TKML administrations the leader or trainee first reads one of the scenarios, then rates on a nine-point scale the appropriateness of each of six to ten or more alternative actions that are described. This is not a multiple-choice test in the sense that only one answer is correct. The alternative actions are all acceptable or unacceptable to various degrees, the mix varying from scenario to scenario, and the "right" rating of each alternative defined by expert consensus. The test was first validated as a survey over three levels of leadership: platoon, company, and battalion command**.** As expected, more experienced leaders agreed on the most effective courses of action to a much greater extent (Hedlund et al., 2000) than less experienced officers, reflecting tacit leadership knowledge acquired on the job.

An example of one of the platoon-level scenarios is:

You are a new platoon leader who takes charge of your platoon when they return from a lengthy combat deployment. All members of the platoon are war veterans, but you did not serve in the conflict. In addition, you failed to graduate from Ranger School. You are concerned about building credibility with your soldiers. What should you do?

Rather than have respondents rate pre-generated alternatives as usual, after reading the scenario, we had officers write open-ended responses as to the course of action they would take. They were given instructions orally as part of their introduction to the threaded discussion environment.

Online Discussion Group Environment—Knowledge Post (KP)

Knowledge Post is a standard threaded discussion environment that has been enhanced with LSA. Its main features are as follows:

1. Users read scenarios or discussion prompts stored as notes within the discussion forum.
2. Users write notes in response to those scenarios or prompts.
3. Users can read and respond to the notes of other users.
4. Users are provided links to notes that are semantically similar to theirs or others.
5. Users can have their own responses evaluated by the *Intelligent Essay Assessor* (Foltz et al., 1999).

Features 1-3 are commonly available in other threaded discussion groups. Features 4 and 5 however are unique to *Knowledge Post* and rely on LSA.

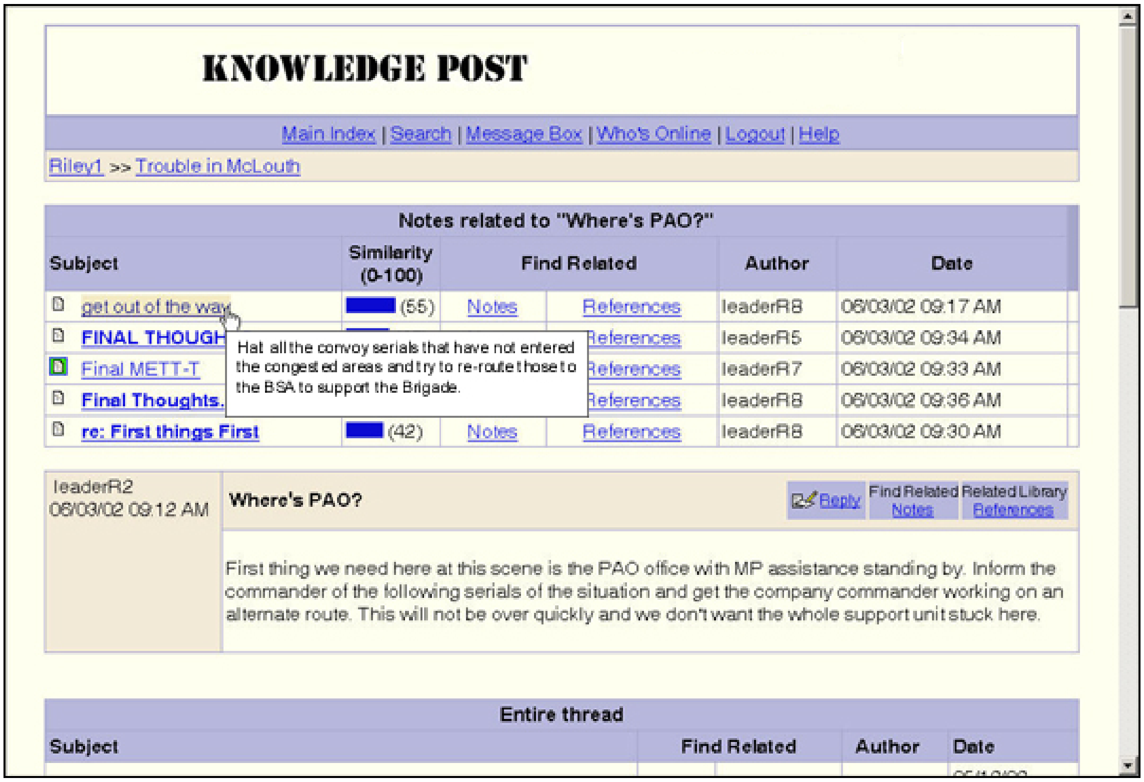


Figure 1. Notes semantically related to leaderR2's note entitled, "Where's PAO?". A PAO is a public affairs officer.

Semantically Related Notes: A Unique Feature of Knowledge Post

In many discussion boards, there are simply too many notes for a participant to read every one. LSA provides a remedy to this situation by allowing users to find contributions that are similar to their own. It thus provides a semantic path through the discussion board. The LSA enabled "Find Related Notes" function in Knowledge Post provides officers with a tool that allows them not only to read responses to their entries, but also read selectively other entries that are not only similar in meaning, but that may introduce related issues. Figure 1 shows the output of "Find Related Notes". The related notes are ordered in the display by their semantic similarity as shown as a number between 0 and 100 (cosine value multiplied by 100). Typically, similar notes have cosines that range from 0.5 to 0.8, overlapping the range of effective values found in Wolfe et al.

The semantic space that LSA creates depends critically on the text corpus over which it is computed. In Knowledge Post, this text corpus is very large, comprising about 12 million words from selected texts that form a representative sample of what a student finishing high school may have read. In addition, the content of several Army manuals, technical books, and descriptions of common Army scenarios are also incorporated. In the past, LSA has shown a robust representation of the meaning of words and passages using similar spaces.

The assumptions underlying LSA propose that similarities and differences in the meanings of words can be largely induced from the contexts in which they appear, and that similarities and differences in the meanings of paragraphs can be induced from a combination of the constituent words. The order of these component words is not nearly as important as the particular selection of words. The meaning is induced not just from keywords, or local co-occurrence of words, but by the solution of many simultaneous equations. Each written contribution is given a position in a multidimensional LSA space based on the sum of the vectors of all its constituent word meanings. To determine the similarity between two contributions, we normally use the cosine of the angle between vectors in the semantic space as a similarity metric.

METHOD

Scenarios

Four Tacit Knowledge of Military Leadership scenarios were presented to officers within KP. A separate topic, on unit readiness, provided an open ended discussion of the officers' understanding of US Army readiness for combat, a central concern for all of them in their daily interactions.

Participants

Eight groups of officers at four Army installations participated in the experiment. Each installation sets aside one week a year during which its personnel are available for research, and the experiment was conducted during this period. Data were collected in small groups of 5 to 15 officers, where group members were all of the same rank. A total of 46 soldiers (11 LTs, 12 CPTs, 13 MAJs, and 10 LTCs) participated for the four platoon scenarios, while a different set of 47 Non Commissioned and Commissioned Officers discussed the Unit Readiness scenario.

Procedure

Each session lasted three hours or less. Copies of the instructions and additional assessment instruments were passed out prior to the experiment. Two introductory pages preceded the instrument. The first page described the purpose of the research and the role of the participant, with an informed consent form. The second page asked for optional information including rank, branch, and current duty position. Oral instructions were provided for logging on to the threaded discussion, and how to respond to each scenario. The scenarios were followed in the same order for each group.

RESULTS

Evaluating the quality of the online contributions.

A total of 199 notes was generated by the Officers for the four platoon scenarios. Each note was scored blind by four expert graders based on their best understanding of the ways to deal effectively with the issues in the scenario and also on the standardized answers available from the extensive administration of the multiple forced-choice 9-point scaling versions of the scenarios. Mean intercorrelations among the four raters of their grades over all 199 notes were significant ($p < .01$): 0.52, 0.40, 0.42, and 0.46. The correlations between any two graders ranged from a high of 0.72 to a low of 0.23.

For each contribution, the most similar notes were found using the cosine similarity metric. We compared the rated quality of each contribution and the rated quality of the nearest note to it, as well as computing the correlation between these two measures. The results are shown separately for each platoon scenario in Table 1.

All four scenarios demonstrated the Goldilocks Principle at work-- the quality of the near notes selected automatically by LSA was significantly higher than the quality of the original note used to search (stem note). At the same time, three of the four scenarios showed a significant correlation between the quality of the stem note and the quality of the nearest note selected by LSA.

Scenario	Mean Quality of Stem Note	Mean Quality of Near Notes	Correlation Between Stem and Near Note Quality
Platoon 1 (N = 47)	6.46	7.38**	0.42**
Platoon 2 (N = 57)	6.94	7.37**	0.28**
Platoon 3 (N = 54)	5.60	6.56**	0.25**
Platoon 4 (N = 41)	6.35	6.66**	0.01

** $p < .01$ (Column 3 significance is based on the difference between the Near Notes and Stem Notes)

Table 1. Average quality of original note and near notes and the correlation between the two.

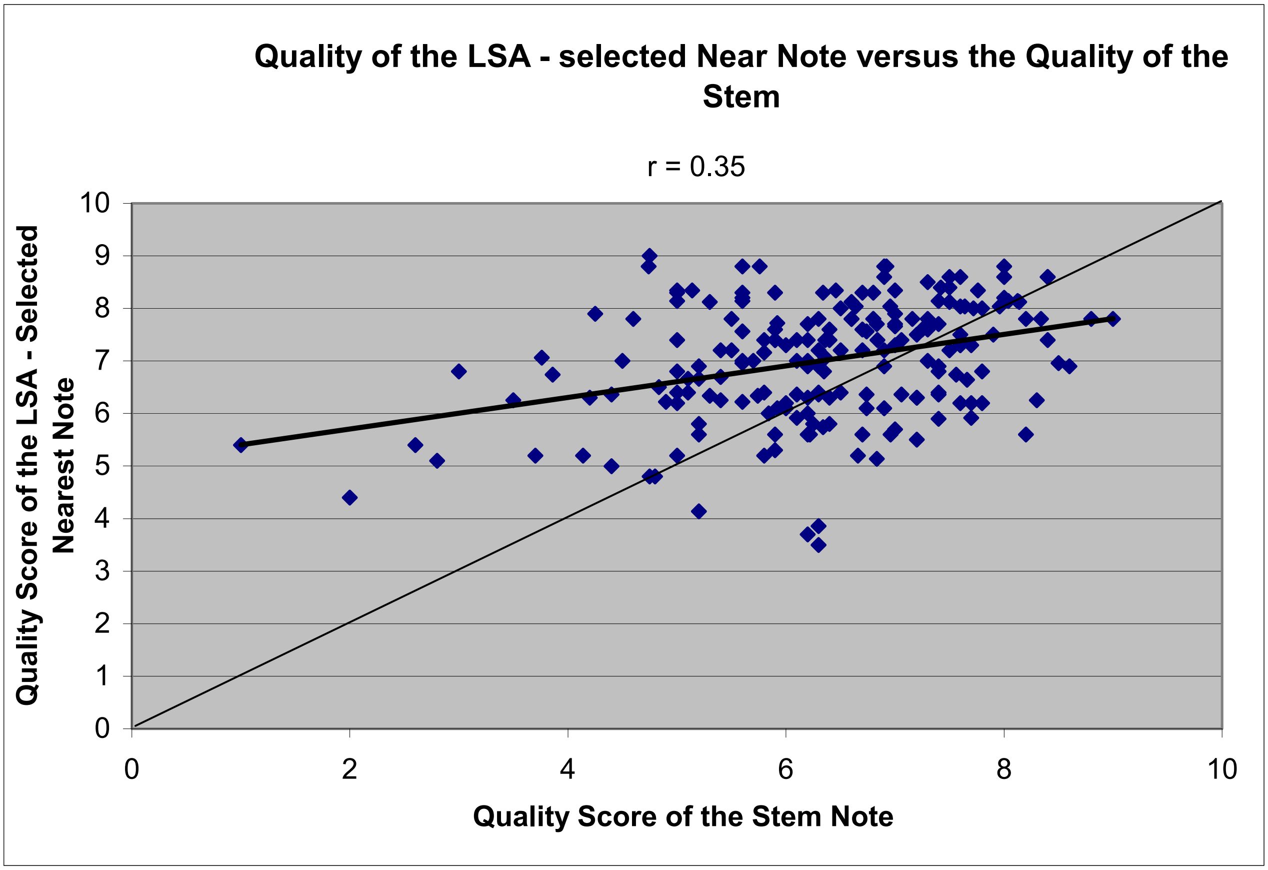
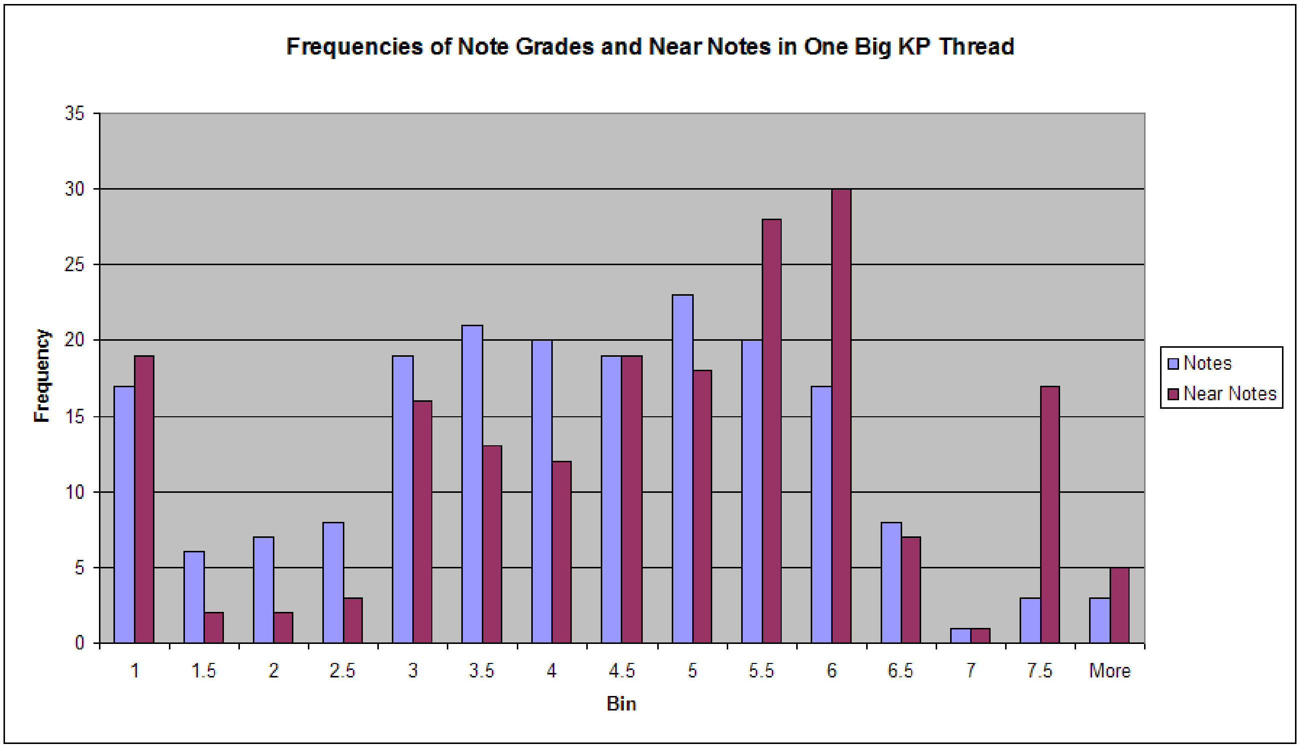


Figure 2. Correlation between the Stem Score and the LSA – selected nearest note score over all 199 notes in the platoon level scenarios.

The overall correlation between the quality of the Stem Note and the quality of the Nearest Note selected by LSA is moderate, but significant ($p < .01$) as shown in Figure 2. However, for the lower half of the Stem Notes, those with a quality score of less then 5, the Near Note is always of higher quality. For those Stem Notes with a quality score greater than 5, the majority of selected Near Notes is also higher in quality.

Figure 3 shows the distribution of the Stem and Near Notes. Of particular interest is the asymmetry of the Near Note distribution, which is markedly skewed to the higher end of the distribution. An exception to this pattern is the lowest bin, but there appears to be an explanation that highlights this exception. These low quality contributions were upon examination, usually irrelevant or extraneous and short comments, outside the knowledge domain; so LSA often linked them to each other but not to the higher quality contributions. These low quality comments are best seen as irrelevant to the issue of how to implement the Goldilocks Principle automatically in LSA.



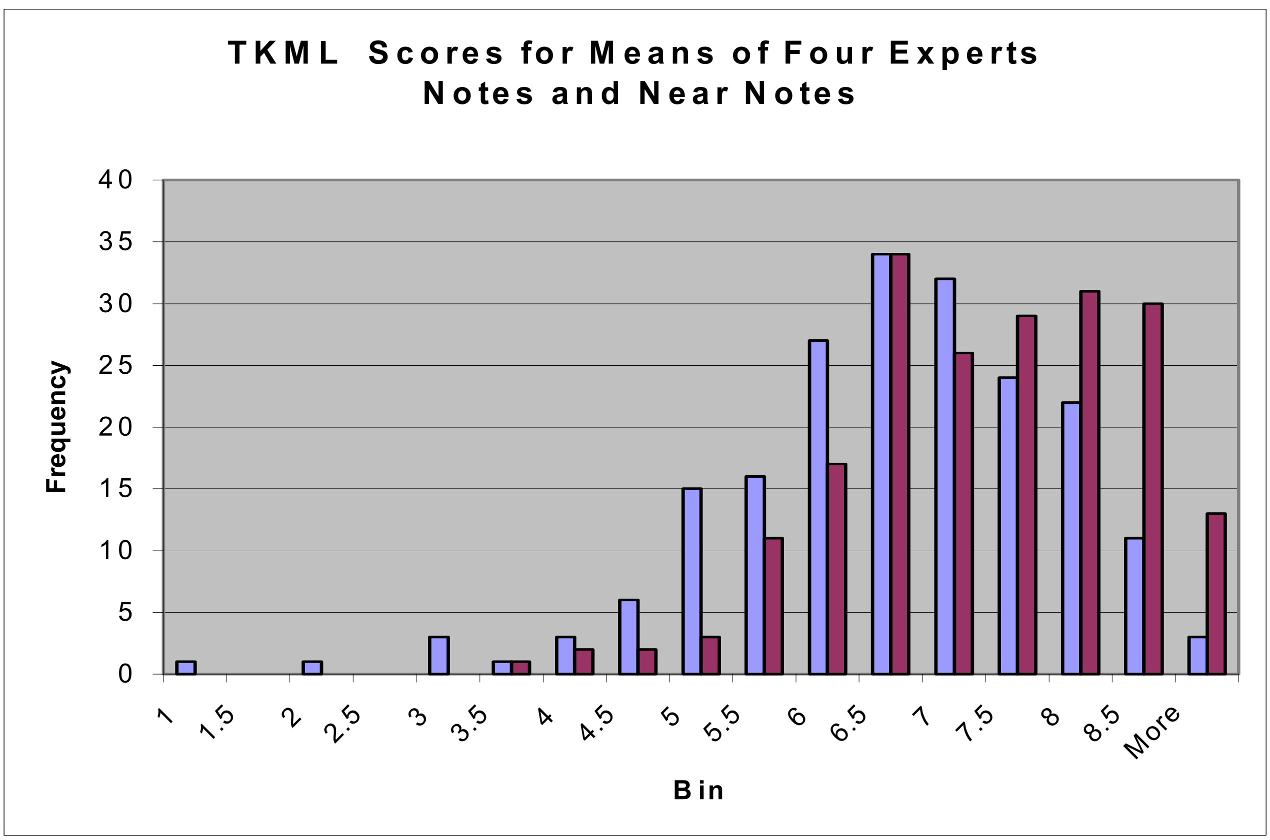


Figure 3. Distribution of mean Near Note grades and Stem Notes for the four Platoon scenarios. Stem notes are shown in blue; Near Notes in maroon.

Unit Readiness Results

The Unit Readiness responses were scored by two military expert graders using their best understanding of the components of readiness in the US Army. Mean intercorrelations between the two raters of their grades over all 192 notes were significant ($r = 0.89$; $p < .01$).

In all, the quality of the Near Notes selected automatically by LSA was significantly higher than the quality of the Stem Note initiating the search ($t = 5.2$; $p < .001$). At the same time there was a significant correlation between the quality of the Stem Note and the quality of the Nearest Note selected by LSA ($r = 0.56$; $p < .01$; 190 df).

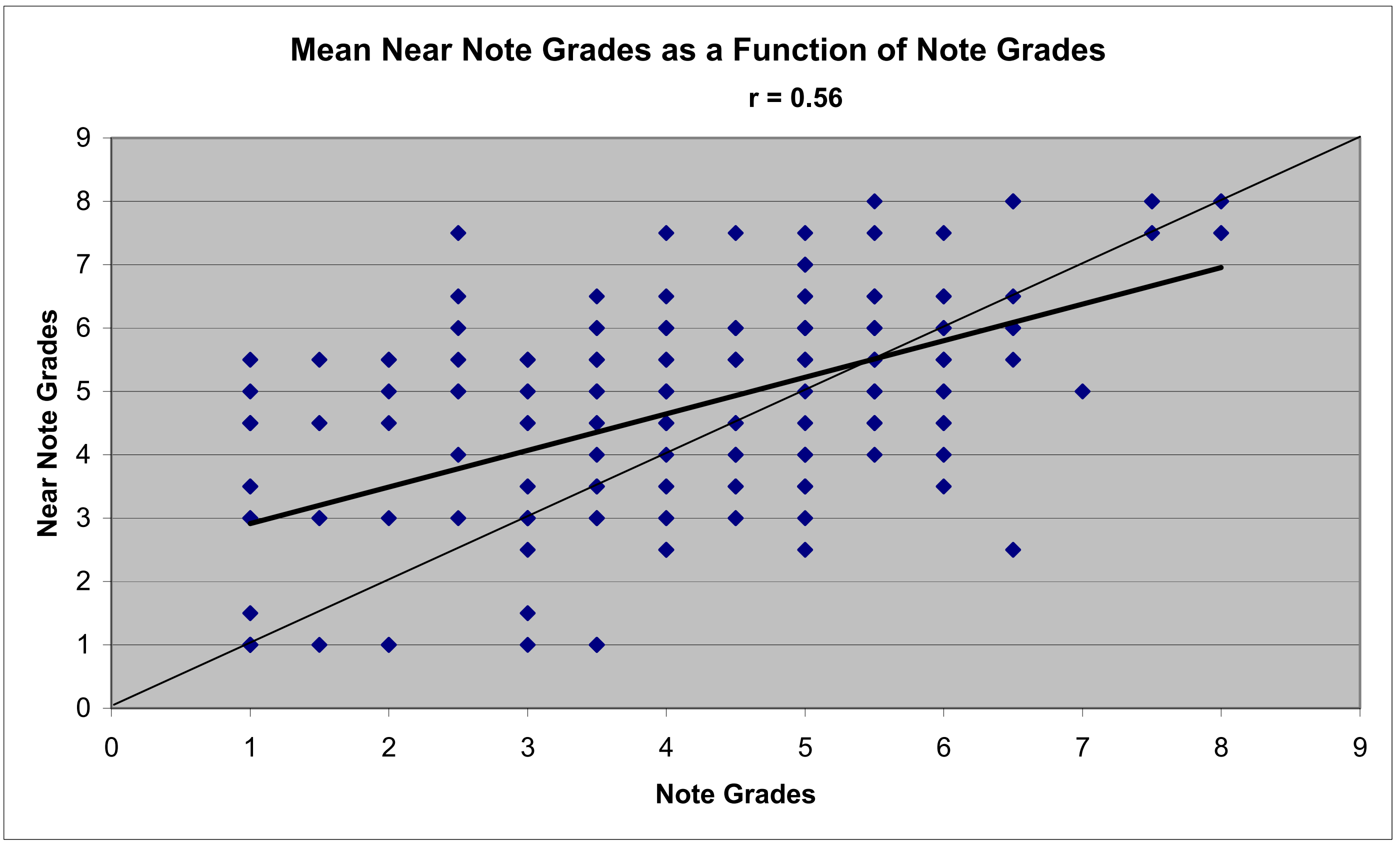


Figure 4. Mean Near Note grades as a function of the related note grades for Readiness Scores.

Comparing the distribution of the notes' grades against the related Near Note grades, the frequencies of the Near Note grades are relatively higher for the higher grades (cf. Figure 5), but the pattern of asymmetrical document-to-document cosines is not nearly as marked as it was for the platoon scenarios. Once again the very low quality notes appear to be selected too often. However, as in the TKML scenarios, these low quality notes are often asides and personal remarks that really should be excluded from the distribution, for the purposes of these analyses.

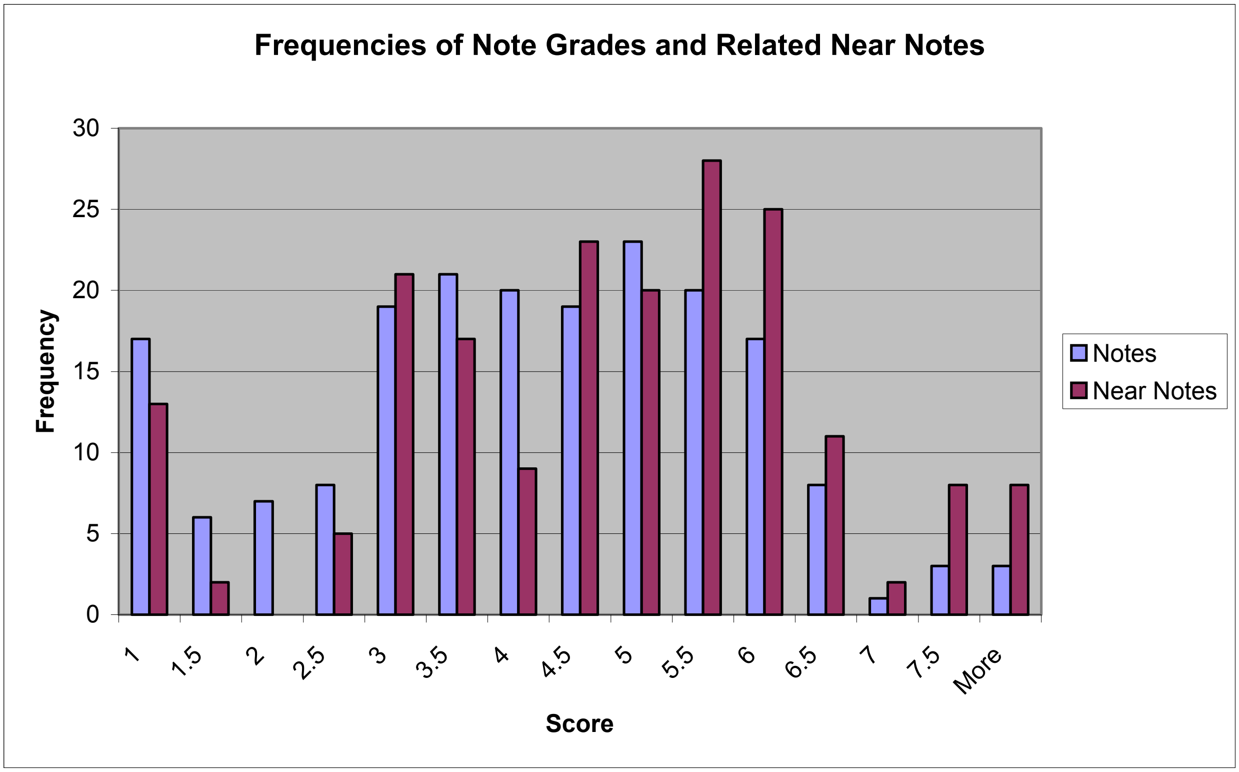


Figure 5. Frequencies of Note Grades compared to the frequencies of Near Note Grades for Readiness Scores.

ANALYSES

On average, in Latent Semantic Space a Near Note is of slightly higher quality than the corresponding Stem Note. But does *reading* a slightly better note lead to more effective learning as Wolfe et al. found? The evidence with regard to the latter question is probably affirmative, but wasn't directly tested in this study.

We have compared officers discussing these same scenarios either face-to-face or using our electronic discussion group (Lochbaum et al., 2002). After reading the scenarios about half of the officers discussed the scenario face-to-face and wrote their responses with pencil and paper, while the other half typed their "solutions" into the online discussion environment. The online group entered an initial response and then a final response after an online synchronous discussion.

Randomized responses were graded blind by two military experts for quality on a 1 to 9 scale with a reliability of 0.78. The results are shown in Figure 6. Those officers who used the online discussion group contributed much higher quality initial responses (shown as First Knowledge Post in the figure below) than the pencil and paper group. Additionally, lower ranking officers learned more using the online discussion group than did the face-to-face participants. How much of this effect can attributed to finding Near Notes while discussing the scenario is not known.

Officers were shown the Near Note facility and used it during the discussions, so some benefits may have accrued. However, this was only one of several factors that may have contributed to higher quality responses. For example, one of the oldest results from the work on electronic chat groups pioneered by Hiltz and Turoff (cf. Benbunan-Fich, Hiltz, & Turoff, 2003) is the greater equality of participation in the electronic medium over face-to-face groups. Anonymity certainly contributes to the increased participation—"nobody knows you're a dog on the internet." In post discussion feedback we have collected on *Knowledge Post*, anonymity is always mentioned as a positive feature, producing richer, more honest communication.

Thus, several factors conspire to produce better discussion and learning in the electronic medium. One is natural desire of humans to communicate with each other. Another is the parallel nature of the discussion—members of an electronic discussion can contribute simultaneously, thereby making more effective use of the time available. This is not possible in face-to-face discussions. The simultaneity of input coupled with greater equality of participation, result in a richer set of ideas generated by a greater number of people. In face-to-face discussions only a few people contribute the bulk of the remarks—in small groups the most vocal two people do over 60% of the talking (Stephan & Mishler, 1952).

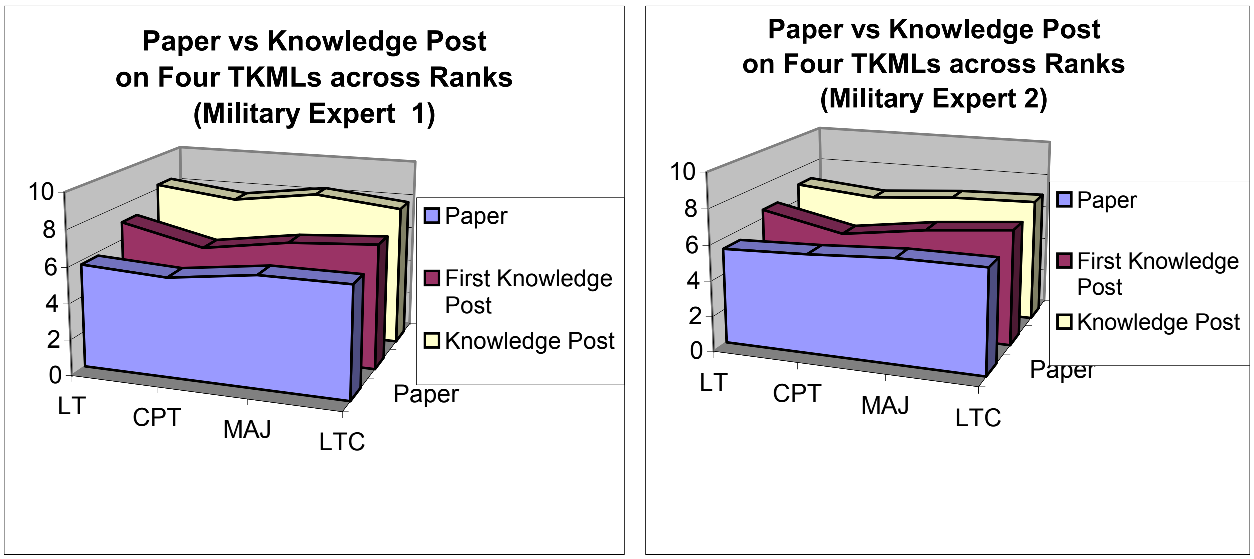


Figure 6. Comparisons of Paper and Pencil to Knowledge Post by Officer Level for the TKML scenarios for two Military Experts' grades (1-9 with 9 being the highest grade).

LSA and Goldilocks

How does a LSA representation of contributions automatically implement the "Goldilocks Principle" in which feedback and additional information are given at nearly the right level of difficulty—the next passage to read being somewhat more advanced than the original? An explanation depends on understanding the properties of the LSA space.

We computed the "centroid" of the high-dimensional vectors all the contributions and then asked whether it was more similar to high quality responses (best one-sixth) or more similar to low quality responses (worst one-sixth). The centroid was much closer to the best responses in terms of cosine similarity (mean cosine between the centroid and the best = 0.74 (SD=.07); mean cosine between the centroid and the worst = 0.60 (SD=.08)). Across all five scenarios the effect was significant ($p < .001$ or less in each case).

Regression to the Mean

Some sort of regression to the mean, either statistical or biased by LSA, can be ruled out as an alternative explanation of this phenomenon since the means of the near note distributions are all significantly higher than the means of all the notes. From the scattergrams in Figures 2 and 4 it appears that the effect is somewhat stronger for the poorer contributions, but this may just be a ceiling effect for the better contributions, since there are a very limited number of even better contributions for the system to choose automatically.

Consensual Centroids of the Semantic Space

Another explanation of this phenomenon would be that the distributions of note to note cosine similarities are highly asymmetric, because there are more ways to be bad than good with poor contributions being more different from each other than good ones. Thus, low scores will be farther from the centroid and closer to better notes than poorer notes. This results in a near neighbor having a higher score on the average.

To expand on this analysis, as Tolstoy (1877) pointed out in the opening lines of Anna Karenina:
"Happy families are all alike; every unhappy family is unhappy in its own way."
This Tolstoy principle might be applied to the consensual assessment of quality by LSA of "happy" responses. High quality responses tend to share many features and converge in meaning. Since the officers each have some expertise that they bring to bear on the problem scenario, the common consensual components tend to be those that offer the best solution to the problem scenario, and responses that include more of these components are in general more expert and of higher quality. As a result, the knowledge space of contributions in the threaded discussion tends to center on the consensually best answers, with worse responses scattered in the periphery of the space.

This explanation depends on the information captured by LSA's centroid of the semantic space of all contributions, and the distribution of responses in this semantic space. Better answers lie nearer the center of LSA's semantic space of these contributions, and these answers are best because they capture most of the consensual common components of experts. In the case of the Army officers, the centroid represents the parts of the solution that have been agreed on by several different people, without those components that are not generally agreed on. In this sense it is better than the average contribution. This effect has been seen in other, but not all, sets of essay scores examined.

We tested this notion directly with the four platoon Tacit Knowledge scenarios by examining the existing 9-point rating alternatives that have been normed for these particular scenarios (Hedlund et al., 2000). These alternatives capture succinctly the common recommendations that officers make about things to do under the circumstances. Some of these alternatives are seen as distinctly better than others. For example, for the scenario in which the new platoon leader failed Ranger School and has to command soldiers who have just returned from combat, the text of two relatively good and bad alternatives are shown below:

"Ask the members of the platoon to share their combat experience: Ask what they learned and how it can help the platoon."
(median 9 "extremely good" on a 9-point scale N=358)

"Announce right up front that you are in charge and the soldiers must accept this fact and treat you with appropriate respect."
(median = 3 "somewhat bad" N=358)

To examine whether the centroid contribution had more elements of good responses, we compared a note constructed of all good alternatives (median rating of at least 8), selected from the forced choice text versions of the TKML, with the centroid as well as with the best and worst one-sixth contributions for each of the platoon scenarios. The results are shown in Table 2. Low quality discussion contributions have significantly less similarity to the centroid than do high quality contributions ($t = 6.84$; 62 d.f.; $p < .01$). The contributions that were combined from the text of good alternatives to the scenarios are closer to the high quality contributions than the low ($t = 4.38$; 62 d.f.; $p < .01$). The magnitude of these cosines is significantly less ($p < .001$) than with the real contributions. Evidently, the actual discussion contributions are more detailed than the distilled ones in the TKML.

Text Items	Proximity to Centroid	Proximity to Combined Good Rating Items
Low Quality Contributions (N=32)	**0.60 ** (SD=.08)**	**0.19 ** (SD=.10)**
High Quality Contributions (N=32)	**0.74 (SD=.07)**	**0.30 (SD=.10)**

Table 2. Cosine similarities (Proximity) to the centroid of high and low quality contributions, as well as similarities to the combined text for good rating, forced choice items across the four platoon scenarios used. Low versus High Quality mean differences are both significant **(p < .01)

As a further test of the hypotheses, the authors created summaries of all the contributions for each of the rank groups ((LTCs, MAJs, CPTs, and LTs) trying to capture all of the best points that came up consensually and repeatedly in the discussion. Figure 7 provides a view of the distribution of all *readiness* comments in a space of cosine distances, with these author's summaries of each echelon's contributions (LTCs, MAJs, CPTs, and LTs) in orange clearly near the center of the distribution.

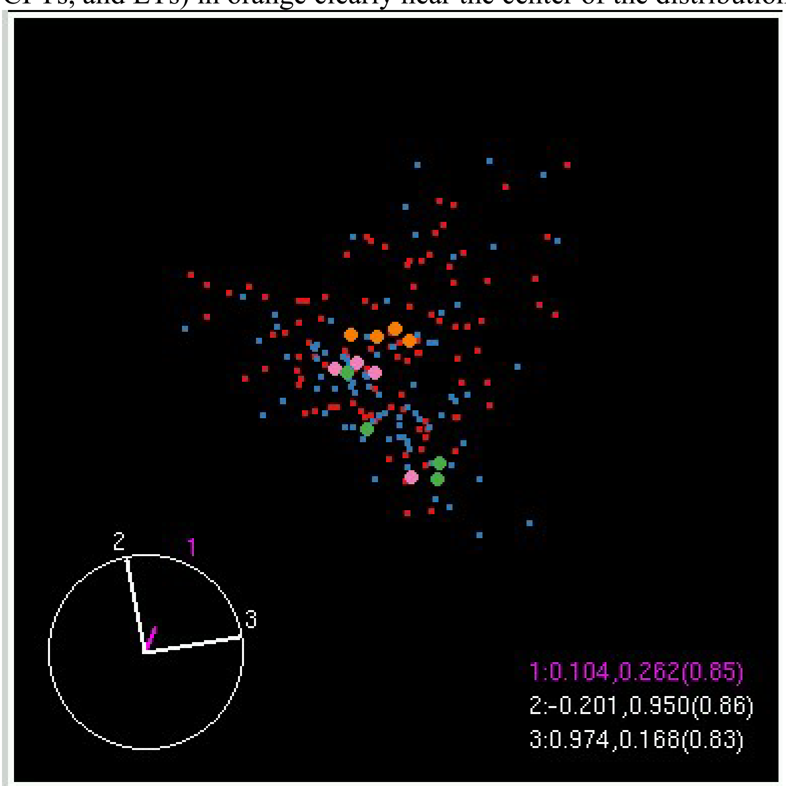


Figure 7. A multidimensional scaling of the cosine distance among readiness responses, with author – generated summaries apparently falling near the center of the space.

DISCUSSION

The Intrinsic Value of Consensus.

Another potential explanation is that notes nearer the center of the space better capture the consensus of the officers, and that the consensus is better not just because the opinions are in fact objectively better, but because consensus itself is valuable. Decision makers need to agree in order to act in concert and to motivate others. The better leaders have learned how to reach consensus more quickly and effectively.

A perhaps somewhat less interesting version of this explanation is that the raters have reached consensus on what is good while reading the notes, and the participants, also being officers, have tended to reach the same consensus. This is less interesting only in the sense that consensus reached by long real-world experience and interaction in leadership situations would probably be more valid than consensus reached through reading discussion notes.

However, the most interesting explanation resides in the possibility that consensus opinions on difficult issues really are better, and somehow closer to the truth, even when there is no absolute or objective standard to make that assessment. Somehow, out of experience, these officers are intuiting parts of better approaches to dealing with complex problems, and the sum of these approaches constitutes something like the best humans can do in these situations. Thus each officer's contribution can be seen as a correlation with this best approach; and their intercorrelations dependent on this correlation with the best approach.

Using LSA to Implement ZPD

The finding that LSA represents a semantic space of contributions to a threaded discussion in which the centroid appears to be composed of better components determined consensually from the space of all answers provided in the discussion provides some insight into how the ZPD can be applied with LSA. By offering a Near Note that is slightly better than each officer's own response, LSA could offer additional (and better) components that are not in the officer's own response, while maintaining some commonality with his or her own response. The officer then has the opportunity of expanding his or her own views to try to encompass the new components, hopefully improving his or her understanding of the issues at the same time. LSA is also sensitive to word grade or difficulty level, so it seems reasonable to assume that LSA or other statistical measure can be used to maintain grade level of vocabulary while introducing more consensual components.

Although we have not tested this hypothesis explicitly, we consider it an interesting direction for further research and we plan to examine other means for testing it empirically.

The way in which LSA models consensus provides an automatic method for selecting the next best piece of material for a student to learn, making this an important contribution to tutoring. Further research is needed to determine more precisely how effective this implementation is in actual learning situations. From the scattergrams, it appears that poor responses elicit near notes that are much better than they are, and this distance may be too great to implement the Goldilocks Principle optimally. For these notes, it may be more effective to have the system automatically select a second or third order near note. Fine tuning of these effects should be relatively easy to implement automatically. However, the evidence from these five case examples is strong that on the whole effective tutoring is occurring automatically in the current system, and provides a first-order implementation of the Goldilocks Principle within the Zone of Proximal Development proposed by Vygotsky (1978).

References

Anderson, J. R. (2002) Spanning seven orders of magnitude: A challenge for cognitive modeling. Cognitive Science, Vol 26(1), Jan-Feb 2002. pp. 85-112. Journal URL: http://www.elsevier.com/inca/publications/store/6/2/0/1/9/4/

Bloom, B. S. (1976). Human characteristics and school learning. New York: McGraw-Hill.

Bruner, J. S. (1966). Toward a theory of instruction. New York: Norton.

Burton, R. R., & Brown, J. S. (1979). An investigation of computer coaching for informal learning activities. International Journal of Man-Machine Studies, 11, 5-24.

Foltz, P. W., Laham, D., Landauer T. K. (1999). The Intelligent Essay Assessor: Applications to Educational Technology. Interactive Multimedia Education Journal. Available online at: (http://imej.wfu.edu/articles/1999/2/04/index.asp).

Hedlund, J., Sternberg, R. J., & Psotka, J. (2000). Tacit knowledge for military leadership: Seeking insight into the acquisition and use of practical knowledge (Tech. Rep. No. ARI TR 1105). Alexandria, VA: U.S. Army Research Institute.

Benbunan-Fich, R., Hiltz, S. R., & Turoff, M. (2003) Decision Support Systems, Vol 34(4), Mar 2003. pp. 457-469.

Landauer, T. K. (2002). On the computational basis of learning and cognition: Arguments from LSA. In Ross, B. H (Ed.) The Psychology of Learning and Motivation, Vol 41, 43-84.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review, 104, 211-240.

Landauer, T. K., Foltz, P. & Laham, D. (1998). An introduction to Latent Semantic Analysis, Discourse Processes. 25, 259-284.

Lochbaum, K., Streeter, L, & Psotka, J. (2002). Exploiting technology to harness the power of peers. Interservice/Industry Training, Simulation and Education Conference, Orlando, FL, December 2-5, 2002.

Palincsar, A. S., & Brown, A. L. (1984). The reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction, 1, 117-175.

Psotka, J., Massey, L. D, & Mutter, S. A. (1988). Intelligent tutoring systems: Lessons learned. Erlbaum, 552 pp.

Rehder, B., Schreiner, M. E., Wolfe, M. B. W., Laham, D., Landauer, T. K., & Kintsch, W. (1998). Using Latent Semantic Analysis to assess knowledge: Some technical considerations. Discourse Processes, Vol 25(2-3.

Stephan, F. F. & Mishler, E. G. (1952). The distribution of participation in small groups: an exponential approximation. American Sociological Review, 17, 1952. pp. 598-608.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., Snook, S. A., & Grigorenko, E. L. (2000). Practical intelligence in everyday life. New York: Cambridge University Press.

Tolstoy Leo (1877) Anna Karenina. Great Literature Online. Http://www.underthesun.cc/

Vygotsky, L. (1978). Mind in Society: The Development of Higher Psychological Processes. Cambridge, MAL: Harvard University Press.

Wolfe, Michael B. W.; Schreiner, M. E.; Rehder, R.;(1998) Learning from text: Matching readers and texts by latent semantic analysis. Discourse Processes, Vol 25(2-3), 1998. pp. 309-336.

Modeling the User for Education, Training, and Performance Aiding

Dr. Robert E. Foster
Director, BioSystems
3080 Defense, Pentagon
Washington, DC 20301-3080

703-588-7437 (voice)
703-588-7545 (fax)
robert.foster@osd.mil

Dr. J. D. Fletcher
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311-1882

703-578-2837 (voice)
703-931-7792 (fax)
fletcher@ida.org

Summary

If we are to present instruction that is available anytime and anywhere, takes advantage of the substantial tutorial efficiencies of one teacher for every student, and is affordable, we must have recourse to technology, specifically computer technology. Such technology can be used in instructional applications that range from drill and practice and tutorial dialogues, to multiplayer simulations and games. It can be used in stand-alone modes or it can be used to supplement classroom instruction. It can be used by individuals or by groups. In all cases, however, it must take account of the current state of the learner, the eventual state of the learner that the instruction is intended to produce, and the instructional techniques that reliably effect transitions from one state to the other. Models of the learner that represent these current and objective states must to an appreciable extent be models of the learner's cognition, which produces the skills, performance, and competence needed for success in all military operations. These models may be implicit as found in intrinsically programmed instruction, or they may be explicit. Both types may be seen in technology-based instruction from its beginning. Early explicit models were largely quantitative, involving fairly simple instructional paradigms, but fairly elaborate mathematics, including instructional applications of optimal control theory. Current efforts are more concerned with qualitative models, 19 of which are briefly described and discussed. These models all contribute to some degree to the efficiency and effectiveness of technology-based instruction. However, new challenges have arisen from today's uncertain, asymmetric operational environment, which may require responses that cannot be foreseen nor well prepared for in advance. Instead we must prepare our military forces and personnel to expect the unexpected and be prepared for it with individual and collective agility, creativity, and adaptability. These qualities are fundamentally cognitive in nature and require more powerful and comprehensive cognitive models if they are to successfully serve our programs of education, training, and performance aiding.

Introduction

This paper concerns research on digital representations of human cognitive processes that may be used to develop computer-mediated learning and performance aiding systems. We refer to such representations as 'models' of human cognition. This topic turns out to be extensive in both breadth and depth, so we focused our discussion on the following three questions:

What is the military value of these models?

What is their current state of development?

What is their relationship to instructional systems development?

What research and development should be undertaken to advance their value and utility?

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training",
held Genoa, Italy13 – 15 October 2003, and published in RTO-MP-HFM-101.

Many valuable contributions are being made by researchers who are modeling neurological activity at one end of the behavioral spectrum and by researchers who are modeling human physical activity and performance at the other. This paper is aimed somewhere in the middle of these efforts. It concerns models, or representations, of human cognitive processes such as perception, memory, learning, decision-making, and problem solving. These processes arise from "micro" neurological activity at one end of the spectrum and, in turn, produce "macro" physical activity and performance at the other. Eventually, research and development may yield models that unify the full spectrum of human behavior from neurons to psychomotor activity, but the current state of knowledge limits us to efforts to understand and model components of this spectrum, hence our focus in this paper on one such component – cognition – which seems an appropriate level of concern for learning and human performance.

Throughout the paper we refer to 'the user'. This is intended as a catchall term for students, decision makers, technicians, analysts, and anyone else who may be using computer technology for education, training, and decision and performance aiding. The paper focuses on digital representation of these users' cognitive processes.

The Military Value of Cognitive Modeling

It is a fact both obvious and frequently neglected that human competence, which is a product of human cognition, is essential to every military operation, across all echelons of command and activity. Its importance is perennially evident in the conduct of military operations. Even in the increasingly technology-saturated environments of modern operations, human competence is needed to launch and control systems in space, operate and maintain robotic vehicles, deploy remote sensors and systems in contested territory, and so forth. In short, there are no unmanned systems. Without competent people to operate, maintain, and deploy our materiel assets, investments in them will return little and may in effect be wasted. Given the wide availability of technology and the ease with which it can be obtained, human competence may increasingly account for the difference between success and failure in military operations. Its availability to commanders anytime and anywhere it is needed is a matter of first importance.

How might we ensure this availability? Training (and its performance aiding analog) provides one means to accomplish this objective, particularly if it can be delivered anytime and anywhere. For example, we might supply each user with an omnipresent tutor. Such tutoring is probably best done by a human who possesses expertise in the relevant subject matter, a comprehensive range of tutorial techniques, and sufficient knowledge of the user to identify, establish, and sustain in that individual the precise human competence needed. Research has shown such tutoring to be extremely effective, producing an often-noted two standard deviations of improvement over less accessible and less effective classroom instruction (Bloom, 1984). Such an approach has been called an instructional imperative and an economic impossibility. It would be maximally effective but remains unaffordable because, for many obvious reasons, we cannot supply every user with a human tutor. This situation creates a gap between what is needed and what we can afford. As in many other endeavors, we are attempting to apply technology to fill this gap.

The research evidence suggests that such applications of technology can succeed. In nearly 300 studies comparing classroom (one teacher, many students) with computer-mediated, individualized instruction (one computer teacher, one student) across many different settings and subject matters, we find a 'rule of thirds' emerging. That is to say that, compared to classroom instruction, technology-based instruction costs about a third less and, additionally, either increases achievement by a third when instructional time is held constant, or decreases time to reach constant levels of achievement by about a third. More detailed discussions of these data have been presented by Fletcher (1997, 2002), Foster and Fletcher (2002), Kulik (1994), Niemiec, Sikorski, & Walberg (1989), and others. The primary payoff for military operations is, of course, the more rapid and reliable preparation of personnel to perform operational duties, producing significant payoffs for resource expenditures, readiness, and, most importantly, operational effectiveness.

Similar research evidence exists in support of technology used to aid performance and decision-making. For instance, technicians with only general training have been found to perform as well as specialists (who required time-consuming and expensive training) if they are provided hand-held or wearable performance aids (e.g., Fletcher and Johnston, 2002; Joyce, 2001; Wisher and Kinkaid, 1989). These aids contribute to military readiness and effectiveness not only by releasing individuals earlier for operational duty, but also, by enhancing human competence for maintaining, operating, and deploying materiel assets -- thereby significantly enhancing materiel readiness.

Costs also matter, of course. For instance, the United States military spends about $4 billion a year on specialized skill training. This is the training provided after "basic" or accession training to qualify personnel for the many technical jobs (e.g., wheeled vehicle mechanics, radar operators, avionics technicians, oceanographers, medical technicians) needed to perform military operations. It does not include aircraft pilot training, field training, or factory training, which are covered in separate cost categories. If the US were to reduce by 30 percent the time to train 20 percent of the personnel undergoing specialized skill training, it would save over $250 million per year. If it were to do so for 60 percent of the personnel undergoing specialized skill training, it would save over $700 million per year (Foster and Fletcher, 2002). These are appreciable savings by most standards.

What do these analyses and observations have to do with cognitive modeling? Effective education, and training must start with a dynamically updateable understanding, or model, of the current state of the user, a model of the knowledge, skills, and abilities the user should attain, and instructional techniques, strategies, and processes for getting from one to the other. This sort of modeling occurs in classroom learning where teachers continually assess what their students know, the level or degree to which they know it, and the most efficient ways to progress in achieving instructional goals. As discussed in the next section, this modeling is also found, both implicitly and explicitly in effective technology-based instruction.

Similar modeling processes are also required to support performance and decision aiding, even though the emphasis in these applications is on problem solution rather than learning. A model of the user is needed to provide advice that can be understood and carried out, as well as a model of the system or situation with which the user is interacting, and the ability to maintain something similar to an instructional dialogue to help the user identify correct solutions or decisions.

In both cases, concern with the knowledge, skills, and abilities that comprise human competence leads us to human cognition and the need to map current cognitive states onto goal cognitive states and determine what must be done next. This presents severe difficulties for classroom instruction (one teacher, many students). For instance, a problem arises from the degree to which students in a typical classroom vary in their prior knowledge, abilities, and learning progress. Research suggests variation by about a factor of five (e.g., Corno & Snow, 1986; Gettinger & White, 1980; Gustaffson & Undheim, 1996; Tobias, 1989). Especially notable for military applications is the observation that variability in prior knowledge increases with age and may be more important in determining progress in such post-secondary venues as military training than it is for students in their earlier years of schooling. Such variability suggests the importance, for both efficiency and effectiveness in military education, training, and performance aiding, of tailoring them to the specific needs of individual users.

Assessment of cognition in classroom instruction is necessarily both informal and imprecise. If we seek to achieve human performance outcomes reliably, anytime, anywhere, and affordably, we must have recourse to technology. If we are to use computer technology to achieve these ends we must be able to represent -- or model in digital form -- current cognitive states, goal states, and ongoing progress from one to the other. The empirical results discussed above, arising from technology-based education, training, and performance aiding, suggest that to some degree we have been successful in doing this. The question then naturally arises as to how well our ability to implement and use such models meets the need for them.

The Current State of Cognitive Modeling

Implicit Cognitive Models

Cognitive models are implemented both implicitly and explicitly in technology-based instruction. Consider the following sample instructional item, which is typical of much, perhaps most, computer-mediated instruction:

In the multiplication 3 x 4 = 12, the number 12 is called a ________.

A.	Factor	{Branch to remedial X1}
B.	Quotient	{Branch to remedial X2}
C.	Product	{Reinforce, go to next}
D.	Power	{Branch to remedial X3}

In this item, the system, the computer instructor, assumes that a student responding "A" misunderstands the meaning of 'Factor,' and lacks an understanding of 'Product', or both. The student will be branched to some instructional materials intended to correct one or the other of these cognitive states and then returned to this item or one similar to it. The same type of remedial approach is applied to responses of "B" and "D". A student responding with "C" may be rewarded, 'reinforced', with encouraging, or positive, feedback and is sent on to whatever item will continue progress toward the instructional goal(s), an action which may by itself constitute positive reinforcement.

The above item appears in an article by Norman Crowder written for Automated Teaching, a book that was published in 1959. We may assume the use of cognitive models is a recent innovation in technology-based instruction, but there is a model of cognition and instructional progress evident in this approach. It covers transitions from unlearned to learned states and illustrates what Crowder called 'intrinsic' programming. This approach stands in contrast to the expensive and difficult-to-prepare 'extrinsic' programming advocated by B. F. Skinner (e.g., 1954), and for good reasons of economy and utility it is the dominant approach (still) in use today in technology-based instruction. It may be found covering many subject matters, posing questions following text paragraphs, graphic displays, simulations, audio presentations, video sequences, and/or other sources of instructional content, but the underlying logic remains the same as Crowder's original – display something, elicit a response, and branch to remedial or reinforcing material depending on the response.

In order to prepare an item such as the above, a developer must both anticipate and prepare responses for several discrete cognitive states, represented by the correct answer (response C) and the 'distractors' (responses A, B, and D) to the item. The cognitive model represented by these states is static, implicit, and limited, but it is there. The main difference between the cognitive modeling in Crowder's (and Skinner's) approaches and the cognitive modeling being developed today is that the earlier models for intrinsic (and extrinsic) programming were implicit, embodied in the program of instruction, whereas today we are attempting to use more explicit models of cognition that we can abstract, express, and validate separately from the systems in which they are used.

Explicit Cognitive Models: Quantitative

These more explicit cognitive models are being used for such applications as intelligent tutoring systems and the human behavior modeling we need to generate computer (automated) military forces for constructive and virtual simulation. Explicit models of cognition were also applied early-on, in the 1960s. They were simple and intended to account for rudimentary learning objectives that could be reduced to something like the substantial amounts of stimulus-response, associative pairing required to learn such material as arithmetic 'facts' (addition, subtraction, multiplication tables), second language vocabulary, and technical jargon (names and functions of biological or mechanical structures). Nonetheless, they led to sophisticated and effective instructional approaches, and the line of research needed to determine the full range of learning situations and objectives to which they could be applied was begun but left unfinished and is rarely found today.

As an example of this approach (and its use of cognitive models) consider the following model of learning (adapted from Paulson, 1973) which attempted to account for the probability that a particular item for a particular learner would transition from the unlearned state (U), to either a short-term learned state (S), i.e., present in working memory, or to a long-term learned state (L), i.e., stored in long-term memory:

		State on Trial n+1				
		L	**S**	**U**		**P(correct)**
State on Trial n	L	**1**	**0**	**0**		**1**
	S	**c**	**1-c**	**0**		**1**
	U	**a**	**b**	**1-a-b**		**g**

In words:

- If a learned item (state L) is presented, then:
- With probability = 1, it stays there.

- If an unlearned item (state U) is presented, then:
- With probability = a, it will transition to long term-memory and the learned state,
- With probability = b, it will transition to a short-term state (S) in working memory from which it can either be learned or forgotten, and finally,
- With probability = 1-a-b, it will remain unlearned.

- If an item is in short-term, working memory (state S), then:
- With probability = c, it will transition to long-term memory and the learned state, otherwise,
- With probability = 1-c, it will remain in the short-term state.

- An item in the short-term state will not slip back to the unlearned state.

This formulation, which is based on Paulson's (1973) discussion, accounts for guessing. As shown in the right-most column above, he assumed a probability = g (presumably for 'guessing') of a correct answer to an unlearned item, but a probability = 1 for a correct answer to an item in the learned or short-term state. The parameters are estimated for each item-student combination.

A key feature of this model is that it accounts for items that are not presented on a trial. In Paulson's formulation -- based on Rumelhart's General Forgetting Theory (1967) -- when an item is not presented, transitions between states are expected to occur in accord with the following transition matrix:

		State on Trial n+1		
		L	S	U
State on Trial n	L	**1**	**0**	**0**
	S	**0**	**1-f**	**f**
	U	**0**	**0**	**1**

In words, when an item is not presented:

• If it is in the learned or unlearned state, it stays there;

• If it is in the short-term state, it may regress to the unlearned state with probability f or remain in the short term state with probability 1-f.

Formulations such as this, which are based on explicit transition models of memory, led to an instructional strategy that has been proven optimal in maximizing the number of items learned in the total time set aside for instruction, T, and allowing for a predetermined number of items, N, to be presented in a single session (e.g., Atkinson and Paulson, 1972). The optimal solution determines which N items to present to a particular student so that the total number of items the student learns is maximized at time T. The solution is roughly the following:

1. Before each trial, identify the item or items in N that have received the fewest number of correct responses since the last error.
2. If only one item is identified, present that item.
3. If more than one item is identified, select from this group the item or items that have been presented the fewest number of times.
4. If only one item remains, present that item.
5. If more than one item remains, select one at random and present it.

This description does not describe how items that have reached criterion in the current pool of N items can be optimally replaced with new items. Such procedures have been discussed by Atkinson and Paulson (1972) and Chant and Atkinson (1973).

Quantitative models of this sort continue to be used in technology-based as particularly evidenced by efforts to apply Bayesian networking to assess the cognitive states of learners (e.g., Van Lehn & Niu, 2001). These models use Bayes' theorem to work backward from users' responses to determine the probabilities that they are using (perhaps have learned) specific cognitive processes. This approach can lead to quite sophisticated models of learners' knowledge and skills.

Three points may be worth making here: (1) Both implicit and explicit models of cognition and cognitive processes have been used in technology-based instruction from its beginning; (2) Fairly simple cognitive models for fairly simple instructional paradigms can lead to sophisticated and effective instructional strategies; and (3) This approach remains a promising line of quantitative research that deserves to be explored more fully.

Explicit Cognitive Models: Qualitative

A line of research and development in cognitive modeling that has been more vigorously pursued in recent years is less quantitative than the above models, but the range of cognition covered tends to be more comprehensive and can thereby be used to meet a wider range of learning objectives. This work typically comes under the heading of 'human behavior modeling' and is increasingly used in the development of simulations for training personnel and units, analyzing tactical, operational, and strategic alternatives, and designing, developing, and acquiring military materiel.

We are fortunate that a number of systematic and comprehensive analyses of these models have recently appeared such as those by Pew and Mavor (1998), who reviewed 11 such models, Ritter, et al. (2002), who reviewed 7 models not covered by Pew and Mavor, and Morrison (2003), who reviewed 19 such models.

The models selected for analysis in these reviews were intentionally devised to be implemented in digital form – in computer algorithms. Doing this for any model is a significant demonstration. If a model can be represented in an algorithm, it can be tested. Using its algorithmic representation to capture and test cognitive processes can significantly enhance both our knowledge of these processes and the effectiveness of our education, training, and performance aiding applications. Diagnostic information indicating where the model is correct, will demonstrate the validity of the model, and indicating where it is incorrect, will suggest where the model must be modified to account for the full range of human cognition. Significant scientific and technological advances can arise from information of this sort, as well as substantial improvements in our ability to educate, train, and assist military personnel.

As Morrison (2003) points out, most of these models are systems of if-then, condition-response ('production') rules that simulate cognitive structures and processes. The 19 models he reviewed, which provide a snapshot of the current state of human cognition and behavior representation, are summarized in Table 1.

Table 1. Summary Descriptions of Cognitive Aspects in Models Reviewed by Morrison (2003)

Model Name	Summary Description	Reference(s)
Atomic Components of Thought (ACT)	Intended to provide a unified theory of mind and a design basis for instructional environments (e.g., intelligent tutors, computer generated forces) and human interfaces. Distinguishes between declarative knowledge (represented with semantic networks) and procedural knowledge (represented using if-then rules).	(Lebriere, 2002) (Anderson, Bothell, Byrne, and Lebriere, 2002)
Adaptive Resonance Theory (ART)	Family of neural net models designed to explain sensory-cognitive processes (e.g., perception, recognition, attention, reinforcement, recall, and working memory). Postulates bottom up (e.g., perceptions) and top down (e.g., expectations, attention control) functions in working memory that interact to produce learning.	(Grossberg, 1976a; 1976b) (Krafft, 2002) (http://web.umr.edu/~tauritzd/art)
Architecture for Procedure Execution (APEX)	Intended to reduce time and effort needed to develop models of human performance in complex, dynamic environments such as simulations, explorations of human performance theories, and assessments of equipment design on human performance. Includes goal-directed action selection for tasks and procedures and resource allocation for perceptual (mostly visual), cognitive, and psychomotor functions.	(Freed, Dahlman, Dalal, and Harris, 2002) (http://www.andrew.cmu.edu/~bj07/apex)
Business Redesign Agent-Based Holistic Modeling System (Brahms)	Models social as well as man-machine interactions. Uses agents to model interactions among physically dispersed groups (e.g., teams), and if-then rules ("detectables" and "beliefs") to model decision making (via "thoughtframes") and behavior within the groups. Emphasizes ethnographic analyses and socio-technical work practices, activities shaped by socio-technical environment, and constructivist, situated cognition to model cognition and behavior.	(Sierhuis and Clancey, 1997) (Clancey, Sachs, Sierhuis, and van Hoof, 1998) (Acquisti, Clancey, van Hoof, Scott, and Sierhuis, 2001)
Cognition and Affect Project (CogAff) {with associated SimAgent toolkit}	Conceptual space for describing cognitive architectures. Integrates emotional with cognitive processes. Incorporates three layers of cognition (reactive, deliberative, and reflective or meta-cognitive), three layers of information processing (perception, central processing, and action), and three types of emotions (primary based on reaction, secondary based on deliberation, and tertiary based on reflection) all producing different perceptual, memory, and motor functions.	(Sloman, 2001; 2003) (http://www.cs.bham.ac.uk/~axs/cogaff.html)

Model Name	Summary Description	Reference(s)
Cognition as a Network Of Tasks (COGNET) {with associated GINA and iGEN™ toolkits)	Intended for cognitive task analysis and description of work domains in multi-task environments requiring contemplative, decision-oriented, open-ended responses. Uses three subsystems to represent information processing (sensory/perceptual, mental modeling, action/motor), four forms of if-then rule-based task knowledge (goal directed task hierarchies, perceptual demons to guide attention, blackboard for organizing declarative information, and possible actions linked to time and resource requirements), and meta-cognitive functions. Allows interfacing with other applications.	(Zachary, Campbell, Laughery, Glenn, and Cannon-Bowers, 2001) (http://www.chiinc.com/cognethome.shtml)
Cognitive Complexity Theory (CCT) {with associated GLEAN3 toolkit}	Focused on human interface design, human-computer interaction, and sequential task performance. Employs device models (transition networks), user models (sequentially executed if-then rules, the fundamental CCT units of cognition, retrieve from long-term memory), and mental operators to represent covert cognitive processes. Long-term memory	(Kieras and Polson, 1985) (Kieras, 1999)
Cognitive Objects within a Graphical EnviroNmentT (COGENT)	Intended solely to provide tools (via a visual programming environment that evolves with the model being built) for cognitive modeling, assuming functional modularity (cognition as interaction among semi-autonomous subsystems) and using low-level processing components.	(Cooper, Yule, and Sutton, 1998) (Yule and Cooper, 2000) (http://cogent.psyc.bbk.ac.uk)
Concurrent Activation-Based Production System (CAPS)	Hybrid model for central cognitive functions (e.g., reading comprehension). Primary focus is on modeling patterns of brain activation patterns in high-level cognition via if-then rules for specific areas of the brain and associative networks for cognitive subsystems. Total activation in working memory is capped, concerned exclusively with declarative knowledge (facts), but with different limits for different individuals. Long-term memory includes procedural and declarative knowledge.	(Just, Carpenter, and Varma, 1999) (http://coglab.psy.cmu.edu/projects_set.html)
Construction-Integration Theory (C-I Theory)	Uses a symbolic theory of sentence comprehension and propositions (actions and objects of the action) stressing goal formation to provide a general model of cognition. Comprehension progresses from approximations to verified integration through mutually reinforced associations and spreading activation in memory. Extended to cover comprehension of novel computer interfaces (LInked model) and new websites (CoLiDeS model) and to incorporate concepts from Latent Semantic	(Kintsch, 1998) (Landauer & Dumais, 1997) (Kitajima & Polson, 1997) (Kitajima, Blackmon, & Polson, 2000) (http://psych-www.colorado.edu/ics)

Model Name	Summary Description	Reference(s)
	Analysis (LSA) used to derive meaning from text.	
Distributed Cognition (DCOG)	Intended to model individuals' expert behavior with agents that use multiple strategies to respond to a complex environment (air-traffic control). Based on a two dimensional space: Abstraction with three levels (skill-based responses to signals, rule-based responses to signs, and knowledge-based responses to symbols) and Decomposition (ranging from individual component to total system processing). Processing within this space depends on level of expertise, workload environment, and an individual's preferred level of engagement.	(Eggleston, Young, & McCreight, 2000) (Eggleston, Young, and McCreight, 2001)
Executive Process/Interactive Control (EPIC)	Intended to model details of peripheral cognitive processes, input (perception) and output (psychomotor responses) to inform human-system interface design by predicting the order and timing of responses. Includes long-term storage of declarative and procedural knowledge and working memory for assessing their application. Capacity and retrieval limitations arise only from perceptual and/or psychomotor systems, not from central memory store.	(Kieras & Meyer, 1995) (http://www.eecs.umich.edu/~kieras/epic.html)
Human Operator Simulator (HOS)	Intended to inform human-system interface design by modeling human performance based on the sequence and timing of subtasks organized in networks. Uses simulation objects (configuration of displays and controls), task networks (if-then rules selecting verb-object pairs used to manipulate the objects), and micro-models (times to complete required subtasks involving perception, information processing, and psychomotor responses) to determine human response times.	(Wherry, 1976) (Harris, Iavecchia, & Dick, 1989) (Glenn, Schwartz, & Ross, 1992)
Man-machine Integrated Design and Analysis System (MIDAS)	Intended to inform human-system interface design by modeling individuals and interactions among individuals in performing multiple, concurrent tasks. Uses sensory input (operators and perceivable – detectable, recognizable, and identifiable – objects), memory (with declarative – beliefs in long-term memory, contexts in working memory – and procedural components), decision-making, attention (with limitations on processing resources), situation awareness (actual and perceived), and psychomotor output to model human operator limitations and capabilities.	(Corker & Smith, 1993) (Hart, Dahn, Atencio, & Dalal, 2001) (http://caffeine.arc.nasa.gov/midas/index.html)

Model Name	Summary Description	Reference(s)
Micro Systems Analysis Of Integrated Network Of Tasks (Micro Saint) {May include the Integrated Performance Modeling Environment (IPME), using HOS micro-models, and WinCrew for estimating workload}	Simulation tool that uses a detailed task analysis to decompose human performance into a networked hierarchy (with branching logic and sequential dependencies) of discrete tasks and subtasks for which performance estimates can be validated. Network consists of subtask nodes (with launching conditions, time to complete, and effects) and relationships (that may be probabilistic, tactical requiring a threshold value, or multiple initiating more than one subtask). Designed to communicate with other models and applications through middleware.	(Laughery & Corker, 1997)
Operator Model Architecture (OMAR) {Uses Developers Interface, a graphics toolkit, for developing performance models.}	Models human behavior as interactions among independent computational agents representing interacting individuals or cognitive processes within individuals. Allows both sequentially dependent and parallel task performance with order determined by activation levels of tasks – without an explicit executive process. Allows facile interface with other models.	(Deutsch, MacMillan, & Cramer, 1993) (Deutsch, 1998) (Cramer, 1998)
PSI (Not an acronym)	Attempts to integrate motivation with cognitive processes. Based on three levels of needs that interact to determine motive strength and specific goal behaviors: System needs (water and energy), Preservation level (pain avoidance), Information level (certainty, competence, affiliation). Action strategies first seek automatized skills, then knowledge-based behavior, then trial and error to satisfy goals.	(Bartl & Dörner, 1998) (Ritter, et al., 2002) (http://www.uni-bamberg.de/~ba2dp1/psi.html)
Situation Awareness Model for Pilot-in-the-Loop Evaluation (SAMPLE)	Generalized from original effort to model situation awareness of pilots and air crews in air combat. Uses cognitive task analyses, pattern recognition from Klein's Recognition-Primed Decision-Making, Endsley's three levels of awareness (detection, identification, and prediction), and Rasmussen's three tiers of action strategy (skill-based pattern recognition, standardized if-then rules, and knowledge-based problem solving) to provide three stages of processing: information processing (with a continuous state estimator and a discrete event detector), situation assessment (with the information fusion and reasoning required by multi-tasking), and decision-making (with a procedure selector and a procedure executor). Output includes information disparity, situation awareness disparity, and combat advantage index.	(Rasmussen, 1983) (Endsley, 1988) (Klein, 1989) (Mulgund, Harper, & Zacharias, 2002)

Model Name	Summary Description	Reference(s)
State, Operator, And Result (SOAR)	Intended as a comprehensive model of human cognition focused on operational task domains depicting all behavior as goal-driven movement through problem spaces that define states and operators for the task(s) at hand. Uses a four-cycle iterative process involving: Input (via human perception), Elaboration (matches if-then, condition-action rules in long term memory with those in working memory to issue proposals for decision making and direct commands for psychomotor actions), Output (psychomotor execution), Decision (either selects operators or identifies 'impasses' requiring a new subgoal until all impasses are resolved). Uses a single process for long-term memory, learning, task representation, and decision-making. All learning occurs through "chunking," which occurs through impasse subgoaling and resolution. Emotions arise from situation awareness clarity and confusion. Integrates individual and team knowledge and allows goals and plans to be shared among team members.	(Lewis, 2001) (http://ai.eecs.umich.edu/soar) (http://www-2.cs.cmu.edu/afs/cs/project/soar/public/www/home-page.html) (http://www.isi.edu/soar/soar-homepage.html) (http://www.nottingham.ac.uk/pub/soar/nottingham/soar-faq.html) (http://phoenix.herts.ac.uk/~rmy/cogarch.seminar/soar.html)

How might these models contribute to the development of computer-mediated learning and performance aiding environments? As suggested above, a model intended to support education and training needs either an implicit or explicit model of cognition if it is to assess the state of a learner's knowledge, skill, and abilities. To do this, it must represent memory and its interactions with other cognitive functions such as perception and attention. It may also represent such cognitive functions as decision-making and problem solving as well as cognitive responses to the environment such as social behavior and situation awareness and/or the extent of cognitive workload.

However, if a model is to support education and training, it is not enough for it just to represent the current state of cognitive processing. It must also represent and project its evolution and development. In short, it must include a model of human learning. Table 2, taken directly from Morrison, summarizes the cognitive functions covered by the models summarized in Table 1. It indicates which models explicitly represent one or more of the following cognitive processes: perception, psychomotor performance, attention, situation awareness, short-term memory, long-term memory, learning, decision-making, problem solving, cognitive workload, emotional behavior, and social behavior.

The table indicates that:

- All 19 models represent decision making – but it is largely the reactive form of decision making that is captured in if-then rules.

- All 19 models represent either short- or long-term memory.

- Perception and attention were well represented in 16 of the reviewed models.

- Although only 4 of the models explicitly represented situation awareness, the functions of situation awareness were present in those representing perception and attention.

- Social behavior was represented in only 5 of the models.

- Emotional behavior was represented in only 3 of the models.

- Learning was represented in only 5 of the models as was problem solving. Morrison suggests that this limited representation may be due to the nature of condition-response production models, which can react to the situations contained in anticipated if-states, but which may not adapt well, if at all, to the unanticipated states and conditions that must be accommodated in learning and problem solving.

Table 2. Cognitive and Behavioral Functions Represented in Models Reviewed by Morrison (2003)

Acronym/ Abbreviation	**Cognitive Function Represented**											
	Perception	**Psychomotor Performance**	**Attention**	**Situation Awareness**	**Working Memory**	**Long-term Memory**	**Learning**	**Decision Making**	**Problem Solving**	**Cognitive Workload**	**Emotional Behavior**	**Social Behavior**
ACT	X	X	X		X	X	X	X	X			
ART	X		X		X	X	X	X				
APEX	X	X				X		X				
Brahms	X	X				X		X				X
CogAff	X	X			X	X		X			X	
COGNET	X	X	X	X	X			X	X	X		
CCT	X	X			X	X		X				
COGENT					X	X	X	X				
CAPS			X		X	X		X	X			
C-I Theory			X		X	X		X				
DCOG	X		X		X	X		X		X		
EPIC	X	X			X	X		X				
HOS	X	X	X		X			X				
MIDAS	X	X	X	X	X	X		X				X
Micro Saint	X		X			X		X		X		
OMAR	X		X			X		X				X
Psi	X	X	X		X	X	X	X	X		X	
SAMPLE	X		X	X		X		X				X
Soar	X	X	X	X	X	X	X	X	X		X	X

Note: An "X" entry indicates that the function is represented by the model.

The five models judged to represent learning are: ACT, COGENT, CAPS, PSI, and Soar. All five of these models also represent long-term memory, working memory, and decision-making. All except COGENT also represent perception, psychomotor performance, and attention.

A model of cognition that includes learning is necessary for education and training applications, but it is not sufficient. A model of learning is not a model of instruction. All 19 models, as good as many of them are, lack this component. This component is needed to suggest links between specific instructional interventions and specific learning outcomes – teaching strategies that reliably bring about transitions from the learner's current cognitive state to one capable of producing the intended instructional outcomes.

Instructional Systems Development

Attaining a "model of instruction" centered around models of human cognition would lead to what might be called "engineering of instruction" --instruction viewed as neither art nor science, but as a way to reliably and efficiently produce specified instructional outcomes. Such a capability for development of instructional and performance aiding systems should be based on empirically derived principles that can be realistically applied. Outcomes might consist of general objectives such as ability to transfer knowledge, long-term retention of knowledge and skill, motivation to continue learning, speed of response, accuracy of response, and so forth. The outcomes might be associated with more specific training objectives such as the ability to locate single component failures in the XYZ power supply, pack a reserve parachute, or devise tactical plans.

Fragments of such a capability for engineering instruction have been identified in research literature, data, and findings. Work is needed to organize, substantially expand, and include them as principles to be incorporated in our current models of cognition. In addition, engineering of instruction requires, as an essential foundational element, robust human cognitive models in order for the training, education or performance aiding system to "know" the user and to dynamically adapt to the user's state.

What Research and Development Do We Need?

This brief review of cognitive models applied to automated instructional and performance aiding systems suggests that much good progress has been made but that much remains to be done. We do not yet have the models we need to fully support the broad range of human behavior required for simulations we now use in training, analysis, and acquisition. More generally, we still lack the comprehensive models we need to represent subject matter expertise, levels of student learning, and most especially the links between specific instructional interventions and the development of specifically targeted cognitive abilities needed for competent performance. What research and development should we pursue to achieve short- mid- and long-term enhancements in the state of the art?

This issue was addressed in a workshop held in November 1999 to assess research and development needed to support the Department of Defense Advanced Distributed Learning initiative (Final Report, 1999), in a series of workshops sponsored in 2002-2003 by the Learning Federation (Learning Federation, 2003), and in another HFM Symposium(Foster and Fletcher, 2002). All three sources cover a wide range of issues and organize their results in different categories, but some common findings, specifically concerned with research necessary for the development of cognitive models, emerge from them. These findings, concerning cognitive modeling, are summarized in Table 3 as issues along with some specific research needed to meet these goals and fill gaps in our current capabilities.

Table 3. Issues and Research Requirements for the Development of Cognitive Modeling Summarized from Assessments of Learning Technology Needs

Issue	Research Requirements
Cognitive Theory	• Representation of 'higher order' cognitive capabilities (e.g., decision-making, problem-solving, meta-cognition, pattern recognition, critical thinking, situational awareness, teamwork). • New concepts and theories of cognition and cognitive workload based on new measurement capabilities. • Valid and verified representation of expertise and its development in complex, ill-structured environments. • Knowledge representations and ontologies that allow interoperability and logical operations within and across disciplines.
Human Behavior Representation	• Comprehensive and accurate representation of individual and crew, team, and unit expertise, capabilities, and performance. • Free, cognitively transparent exchange of virtual (avatar) and actual users in crew, group, team learning
Cognitive Model Authoring	• Automated development, verification, and validation of cognitive models. • Automated processes for performing cognitive analysis and cognitive readiness assessment. • Automated capture of expertise -- self-generating, self-modifying data bases built from cases and examples of successful problem solving and decision-making. • Principles for developing physically and cognitively realistic avatars.
User Assessment and Representation	• New forms of computer-administered assessment items using the full display, timing, and natural language understanding capabilities of technology. • Generation of valid, unobtrusive near real time assessment from interactions of individuals, teams, crews, and units with the learning or performance aiding environment. • Representation of subject matter misunderstandings and their sources. • Generation and use of questions to build cognitive profiles of users. • Assessment of cognitive workload.

Issue	Research Requirements
	• Assessment of high-level cognitive skills needed to deal with unanticipated and unexpected situations.
Management of Progress	• Ability to match instructional or problem solving goals with current state of the user and generate or select optimal tutorial and/or problem solving strategies. • Automated principles of design and presentation needed to ensure reliable achievement of targeted cognitive state(s) by individuals, crews, teams, and units. • Automated principles for the development of higher-level cognitive skills such as creativity, adaptability, problem solving, and situation awareness. • Comprehensive understanding of meta-cognition and its development. • Comprehensive understanding of incentive management and its interaction with cognitive development. • Technology-based tools allowing distributed users to manage their own progress and problem solving. • Predictions of learning rate and success from user profile information.
User Interface	• Management of user dialogue based on model of user cognitive abilities, style, and progress toward objective(s)

The efforts suggested by Table 3 are realistic in that that they are amenable to research that can be performed with approaches available from our current state of knowledge. They suggest goals that can be achieved to an appreciable degree in the next 3-5 years. Doing so will be worth the effort and will return much more to the success of our operational capabilities than it will cost.

The value of cognitive models has another, increasingly important dimension. The current world environment presents significant challenges to our capabilities for preparing military personnel to meet them, and thereby to our capabilities for providing military education and training. We have responded in ways that have proven successful in the past, with task lists, essential task lists, mission essential task lists, and even joint mission essential task lists. These task lists suggest education and training objectives that we know how to meet.

However, the current asymmetric, unpredictable operational environment now facing our military personnel will inevitably present situations that are unexpected and for which they may be little prepared. Our people and their allies will have to respond to these situations with agility, flexibility, creativity, and skillful leadership. Their readiness to acquire the additional capabilities needed to meet these unexpected, unforeseen challenges will contribute substantially to the success of their operations. How, then, can we best prepare our people to expect the unexpected and deal with it successfully? Such an aspect of readiness is a cognitive capability. It places special demands on our ability to model cognition and to train both individuals and units. It is an essential component of what we have called cognitive readiness (Etter, Foster, & Steele, 2000), and a combination of technology-based education, training, and performance aiding is expected to help our forces achieve it.

The components of cognitive readiness cover issues that include the following:

Situation awareness, which is generally defined as the ability to perceive oneself in relation to the enemy and the environment. Situation awareness has been shown to improve with practice and instructional feedback.

Memory, which is described as an active, reconstructive process supported by two underlying theoretical mechanisms: encoding specificity, which stresses the importance of external and internal cues, and transfer-appropriate processing, which stresses actions performed during encoding and retrieval. Trade-offs exist between instruction used to enhance retention and speed of initial acquisition. Conditions of learning, particularly those providing overlearning, can be designed to enhance retention.

Transfer of training, which is described as the ability to apply what is learned in one performance context to another. Massive amounts of practice with feedback will enhance "low-road" transfer requiring little cognitive mediation. Training in forming mindful, conscious abstraction will enhance "high road" transfer, which requires cognitive mediation.

Metacognition, which refers to the executive functions of thought, particularly those pertaining to knowledge and regulation of one's cognitive processes and progress toward accepted goals. Metacognitive skills can be enhanced by exercises designed to increase awareness of self-regulatory processes.

Automaticity, which refers to processes that are performed rapidly, requiring few attentional resources. Practice with feedback and overlearning can produce automatic processing in many tasks.

Problem Solving, which transforms goals and subgoals into a plan of action by processes such as trial-and-error, proximity, fractionation, and knowledge-based referrals. Techniques for problem solving matched to goal and situation categories can be successfully taught, as can the information base needed for "strong" problem solving methods, which depend on acquired knowledge.

Decision-Making, which is described as the selection of tactical and strategic plans, which are frequently primed by the recognition of learned patterns. Formal instruction in decision-making techniques may improve the quality of decisions, but some aspects of successful decision-making are determined by individual dispositions.

Mental Flexibility and Creativity, which may be cast as problem-solving, applying "strong" methods based on acquired knowledge and skills, and "weak" methods, used for poorly defined, ill-structured, chaotic tasks. Creativity may be more closely associated with the latter "weak" methods. The research is unclear whether these weak methods can be trained directly. It seems more likely that the facility with which people apply appropriate weak methods (i.e., achieve "creative solutions") to novel situations is determined by native abilities.

Leadership, appears to consist of motivational patterns and a combination of technical, conceptual, and interpersonal skills, the last being the most difficult to acquire and measure. However, technical and conceptual skills that are needed by leaders can, to an appreciable extent, be taught. Interpersonal skills and patterns of motivation required for leadership appear to be more dependent on native abilities and are thus more difficult to teach.

Emotion, must be channeled and controlled if military personnel are to perform complex tasks under the stress and confusion that accompany modern military operations. Deeply engaging, sensory immersing simulations provide promise for training warfighters to retain critical pieces of information and to perform under highly stressful conditions.

These issues have been discussed extensively in research literature and their specific relevance to cognitive readiness have been discussed by Morrison and Fletcher (2002). The points to be emphasized here are that (a) assessment and development of the capabilities suggested by these issues will key on the adequacy of the cognitive models on which our education, training, and performance aiding are based and (b) the adequacy of our cognitive modeling is a matter of first importance in the current unpredictable operational environment.

The modeling efforts reviewed in this paper along with similar efforts involving human cognition represent significant opportunities for cooperative research by the NATO community concerned with the human competence that is an essential component of every military operation. We may wish to rise to the opportunities they present.

References

Acquisti, A., Clancey, W. J., van Hoof, R., Scott, M., & Sierhuis, M. (2001, December). Brahms tutorial (Technical Memorandum TM01-0002, Version 0.9.9.4 RFC). Moffett Field, CA: NASA Ames Research Center.

Anderson, J. R., Bothell, D., Byrne, M. D., & Lebiere, C. (2002, September). An integrated theory of the mind. [Manuscript submitted for publication].

Atkinson, R. C., & Paulson, J. A. (1972) An approach to the psychology of instruction, Psychological Bulletin, 78, 49-61.

Bartl, C., & Dörner, D. (1998). Comparing the Behaviour of PSI with Human Behavior in the BioLab Game (Memorandum Number 32). Bamburg, Germany: Universität Bamburg: Lehrstuhl Psychologie II.

Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13, 4-16.

Chant, V. G., & Atkinson, R. C. (1973) Optimal allocation of instructional effort to interrelated learning strands. Journal of Mathematical Psychology, 10, 1-25.

Clancey, W. J., Sachs, P., Sierhuis, M., & van Hoof, R. (1998). Brahms: Simulating practice for work systems design. International Journal of Human-Computer Studies, 49, 831-865.

Cooper, R., Yule, P., & Sutton, D. (1998). COGENT: An environment for the development of cognitive models. In U. Schmid, J. F. Krems, & F. Wysotzki (Eds.), A Cognitive Science Approach to Reasoning, Learning, and Discovery (pp. 55-82). Lengerich, Germany: Pabst Science Publishers.

Corker, K. M., & Smith, B. R. (1993). An architecture and model for cognitive engineering simulation analysis: Application to advanced aviation automation. In Proceedings of the AIAA Computing in Aerospace 9 Conference. Santa Monica, CA: American Institute of Aeronautics and Astronautics.

Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 605–629). New York: Macmillan.

Cramer, N. (1998). Distributed-OMAR: Reconfiguring a Lisp system as a hybrid Lisp/(Java) component. Paper presented at the Lisp Users Group Meeting. Berkeley, CA: Association of Lisp Users.

Crowder, N. A. (1959) Automatic teaching by means of intrinsic programming. In E. Galanter (Ed.), Automatic Teaching: The State of the Art (pp.109-116). New York, NY: John Wiley & Sons.

Deutsch, S. (1998). Multi-disciplinary foundations for multiple-task human performance modeling in OMAR. Paper presented at the Twentieth Annual Meeting of the Cognitive Science Society. Madison, WI: Cognitive Science Society.

Deutsch, S. E., MacMillian, J., Cramer, N. L. (1993). Operator Model Architecture (OMAR) demonstration final report (AL/HR-TR-1996-0161). Wright-Patterson AFB, OH: Armstrong Laboratory, Logistics Research Division.

Eggleston, R. G., Young, M. J., & McCreight, K. L. (2000). Distributed cognition: A new type of human performance model. In M. Freed (Ed.), Simulating human agents: Papers from the 2000 AAAI Fall Symposium (Technical Report FS-00-03) (pp. 8-14). Menlo Park, CA: AAAI Press.

Eggleston, R. G., Young, M. J., & McCreight, K. L. (2001). Modeling human work through distributed cognition. In Proceedings of the 10th Conference on Computer Generated Forces and Behavioral Representation. Orlando, FL: Simulation Interoperability Standards Organization.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In Proceedings of the 32nd Annual Meeting of the Human Factors Society (pp. 97-101). Santa Monica, CA: Human Factors Society.

Etter, D.M., Foster, R.E., & Steele, T.P. (2000). Cognitive readiness and advanced distributed learning. Crosstalk: The Journal of Defense Software Engineering, 13, 5-6.

Final Report: Department of Defense Continuous Learning System for 2012 (1999). Washington, DC: United States Department of Defense, Director of Defense Research and Engineering.

Fletcher, J.D. (1997) What have we learned about computer based instruction in military training? (pp. 169-177). In R.J. Seidel & P.R. Chatelier (Eds.), Virtual reality, training's future? New York, NY: Plenum Publishing.

Fletcher, J. D. (2002) Evidence for Learning from Technology-Assisted Instruction. In H. F. O'Neil Jr. & R. Perez (Eds.) Technology Applications in Education: A Learning View (pp. 79-99). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fletcher, J. D. & Johnston, R. (2002) Effectiveness and Cost Benefits of Computer-Based Aids for Maintenance Operations. Computers in Human Behavior, 18, 717-728.

Foster, R. E., & Fletcher, J. D. Computer-Based Aids for Learning, Job Performance, and Decision Making in Military Applications: Emergent Technology and Challenges (IDA Document D-2786). Alexandria, VA: Institute for Defense Analyses, 2002.

Freed, M., Dahlman, E., Dalal, M., & Harris, R. (2002). Apex reference manual for Apex version 2.2. Moffett Field, CA: NASA ARC.

Foster, R. E., & Fletcher, J. D. Computer-Based Aids for Learning, Job Performance, and Decision Making in Military Applications: Emergent Technology and Challenges (IDA Document D-2786). Alexandria, VA: Institute for Defense Analyses, 2002.

Gettinger, M., & White, M.A.(1980) Evaluating curriculum fit with class ability. Journal of Educational Psychology, 72, 338-344.

Glenn, F., Schwartz, S., & Ross, L. (1992). Development of a Human Operator Simulator Version V (HOS-V): Design and implementation (Research Note 92-PERI-POX). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Grossberg, S. (1976a). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. Biological Cybernetics, 23, 121-134.

Grossberg, S. (1976b). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. Biological Cybernetics, 23, 187-202.

Gustaffson, J. & Undheim, J.O. (1996). Indivdidual differences in cognitive functions. In D. C. Berliner & R. C. Calfee (Eds.), Handbook of educational psychology (pp. 186–242). New York: Macmillan Reference.

Hart, S. G., Dahn, D., Atencio, A., & Dalal, K. M. (2001). Evaluation and Application of MIDAS v2.0 (SAE Technical Paper 2001-01-2648). Warrendale, PA: The Society of Automotive Engineers.

Joyce, D. (2001) Fusing simulation and performance support – the winning combination for improving equipment readiness? In Proceedings of the 2001 Interservice/Industry Training, Simulation and Education Conference, Arlington, VA: National Training systems Association.

Just, M. A., Carpenter, P. A., & Varma, S. (1999). Computational modeling of high-level cognition and brain function. Human Brain Mapping, 8, 128-136.

Kieras, D. E. (1999). A Guide to GOMS Model Usability Evaluation Using GOMSL and GLEAN3 [On-line publication]. Retrieved August 27, 2002 from the University of Michigan's Electrical Engineering and Computer Science Department site: http://www.eecs.umich.edu/people/kieras/GOMS/GOMSL_Guide.pdf.

Kieras, D. E., & Meyer, D. E. (1995). An Overview of the EPIC Architecture for Cognition and Performance with Application to Human-Computer Interaction (EPIC Report No. 5). Ann Arbor, MI: The University of Michigan.

Kieras, D. E., & Polson, P. G. (1985). An approach to the formal analysis of user complexity. International Journal of Man-Machine Studies, 22, 365-394.

Kintsch, W. (1998) Comprehension: A Paradigm for Cognition. New York: Cambridge University Press.

Kitajima, M., Blackmon, M. H., & Polson, P. G. (2000). A comprehension-based model of web navigation and its application to web usability analysis. In S. McDonald, Y. Waern, & G. Cockton (Eds.), People and Computers XIV-Usability or Else! [Proceedings of HCI 2000] (pp. 357-373). Heidelberg, Germany: Springer-Verlag.

Kitajima, M., & Polson, P. G. (1997). A comprehension-based model of exploration. Human-Computer Interaction, 12, 345-389.

Klein, G. A. (1989). Recognition-primed decisions. In W. Rouse (Ed.), Advances in Man-Machine Systems Research, Volume 5 (pp. 47-92). Greenwich, CT: JAI Press.

Krafft, M. F. (2002, October). Adaptive Resonance Theory. Retrieved March 5, 2003 from the University of Zurich, Department of Information Technology Web site: http://www.ifi.unizh.ch/staff/krafft/papers/2001/wayfinding/ html/node97.html

Kulik, J.A. (1994) Meta-analytic studies of findings on computer-based instruction. In E.L. Baker & H.F. O'Neil, Jr. (Eds.) Technology Assessment in Education and Training. Hillsdale, NJ: Lawrence Erlbaum Associates.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's Problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review, 104, 211-240.

Laughery, K. R., Jr., & Corker, K. (1997). Computer modeling and simulation. In G. Salvendy (Ed.), Handbook of Human Factors and Ergonomics (2nd ed.) (pp. 1375-1408). New York: John Wiley & Sons.

Learning Federation: Learning Science & Technology R&D, A Roadmap to the Future of Learning (2003). Washington, DC: Federation of American Scientists.

Lewis, R. L. (2001). Cognitive theory, Soar. In N. J. Smelser & P. B. Baltes (Eds.), International Encyclopedia of the Social and Behavioral Sciences. Amsterdam: Pergamon (Elsevier Science).

Morrison, J. E. (2003) A Review of Computer-Based Human Behavior Representations and Their Relation to Military Simulations (IDA Paper P-XXXX). Alexandria, VA: Institute for Defense Analyses.

Morrison, J. E., & Fletcher, J. D. (2002) Cognitive Readiness (IDA Paper P-3735). Alexandria, VA: Institute for Defense Analyses.

Mulgund, S. S., Harper, K. A., & Zacharias, G. L. (2000). SAMPLE: Situation Awareness Model for Pilot-in-the-Loop Evaluation. In Proceedings of the 9th Conference on Computer Generated Forces and Behavioral Representation. Orlando, FL: Simulation Interoperability Standards Organization.

Niemiec, R. P., Sikorski, M., & Walberg, H. J. (1989) Comparing the cost-effectiveness of tutoring and computer-based instruction. Journal of Educational Computing Research, 5, 395-407.

Paulson, J.A. (1973) An Evaluation of Instructional Strategies in a Simple Learning Situation (Technical Report No. 209). Stanford, CA: Institute for Mathematical Studies in the Social Sciences, Stanford University.

Pew, R. W., & Mavor, A. S. (Eds.) (1998). Modeling Human and Organizational Behavior:Applications to Military Simulations. Washington, DC: National Academy Press.

Rasmussen, J. (1983), Skills, rules and knowledge: Signals, signs and symbols and other distinctions in human performance models. IEEE Transactions On Systems, Man And Cybernetics, SMC-13, 257-266.

Ritter, F. E., Shadbolt, N. R., Elliman, D., Young, R., Gobet, F., & Baxter, G. D. (2002). Techniques for modeling human and organizational behaviour in synthetic environments: A supplementary review. Wright-Patterson Air Force Base, OH: Human Systems Information Analysis Center.

Rumelhart, D. E. (1967) The Effects of Inter-Presentation Intervals on Performance in a Continuous Paired-Associate Task (Technical Report No. 27). Stanford, CA: Institute for Mathematical Studies in the Social Sciences, Stanford University.

Sierhuis, M., & Clancey, W. J. (1997). Knowledge, practice, activities and people. In B. R. Gaines & R. Uthurusamy (Eds.), Artificial Intelligence in Knowledge Management: Papers from the 1997 AAI Spring Symposium (Technical Report SS-97-01) (pp. 142-148). Menlo Park, CA: American Association for Artificial Intelligence.

Skinner, B. F. (1954) The science of learning and the art of teaching. Harvard Education Review, 24, 86-97.

Sloman, A. (2001). Varieties of affect and the CogAff architectural scheme. From Symposium on Emotion, Cognition and Affective Computing, Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB). Brighton, England: University of Sussex.

Sloman, A. (2003). How many separately evolved emotional beasties live within us? In R. Trappl, P. Petta, & S. Payr, (Eds.), Emotions in humans and artifacts. Cambridge, MA: The MIT Press.

Tobias, S. (1989). Another look at research on the adaptation of instruction to student characteristics. Educational Psychologist, 24, 213-227.

VanLehn, K. & Niu, Z. (2001). Bayesian student modeling, user interfaces and feedback: A sensitivity analysis. International Journal of Artificial Intelligence in Education. 12, printed version to appear.

Wherry, R. J. (1976). The Human Operator Simulator - HOS. In T. B. Sheridan & G. Johannsen (Eds.), Monitoring behavior and supervisory control. 283-293. New York, NY: Plenum Press.

Wisher, R. A., & Kincaid, J. P. (1989) Personal Electronic Aid for Maintenance: Final Summary Report (ARI-RR-1516). Alexandria, VA: US Army Research Institute. (DTIC/NTIS ADA 210 348)

Yule, P., & Cooper, R. (2000, August). The COGENT tutorial. Presented at the 22nd Annual Conference of the Cognitive Science Society. Philadelphia, PA.

Zachary, W., Campbell, G. E., Laughery, K. R., Glenn, F., & Cannon-Bowers, J. A. (2001). The application of human modeling technology to the design, evaluation and operation of complex systems. In E. Salas (Ed.), Advances in human performance and cognitive engineering research, Volume 1 (pp. 199-247). New York: JAI Press.

Generalized Operations Simulation Environment For Aircraft Maintenance Training

Terence S. Andre, Lt. Col, PhD
Winston R. Bennett Jr., PhD
Anna R. Castillo, MS
Dale P. McClain, AOS
Mathew D. Purtee, Lt
Brenda M. Wenzel, PhD
Warfighter Training Research Division
Air Force Research Laboratory
6030 South Kent Street
Mesa, Arizona 85212-6061

Mary Graci, MEd
Maricopa Institute for Virtual Reality Technologies
Chandler-Gilbert Community College
7360 East Tahoe Avenue
Mesa, Arizona 85212

Summary

A training need exits for a common, cost-effective virtual reality simulation platform for aircraft maintenance training. The need is driven by a lack of concurrency between equipment and simulators, inaccessibility of equipment and simulators for training, high life-cycle and sustainment costs for training simulators, and the absence of full-scale assets or mock-up devices at training squadrons. In the absence of training assets and devices, resident training has been restricted to traditional methods such as referencing technical orders and academic courseware. An additional need for on-demand training at operational units has contributed to increased costs by allowing maintainers to use operational flight simulators that often require declassification for maintenance training purposes. This paper describes the assessment of the Virtual Environment Safe-for-maintenance Trainer and development of the Generalized Operations Simulation Environment, the next generation virtual reality aircraft maintenance training program.

1 Introduction

The Virtual Environment Safe-for-Maintenance Trainer

The Virtual Environment Safe-for-Maintenance Trainer (VEST) was developed in 1997 as a cooperative effort among the 363rd Training Squadron (TRS), Sheppard Air Force Base (AFB), Texas, the Air Force Research Laboratory (AFRL) through a contract with Command Technologies Incorporated (CTI) at Brooks Air Force Base, Texas, and the Air Education and Training Command (AETC), Randolph AFB, Texas. The trainer was designed to meet three basic needs: (a) train switchology for F-15E-model two seat cockpits, (b) train F-15E-model weapons stations familiarization, and (c) train specific ground safe-for-maintenance tasks on the F-15E model. Full-scale F-15E model aircraft are not available, as training assets, at the 363rd TRS.

VEST is an immersive virtual reality (VR) environment that provides apprentice technicians demonstrations, drills, and checks on performance maintaining the F-15E model aircraft. The value of VEST is its low-cost replacement of the actual aircraft, thus, providing apprentices with training opportunities they would not otherwise have prior to deployment. VEST provides apprentices with contextualized, three-dimensional, interactive experiences with the F-15E model front and rear crew stations, weapons stations, and ground safe-for-maintenance. Apprentices assigned to bases with F15-E model aircraft are required to complete 21 lessons in VEST, after successfully completing baseline training on the F-15C model aircraft.

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held in Genoa, Italy, 13 – 15 October 2003, and published in RTO-MP-HFM-101.

There are two VEST stations in operation at the 363rd TRS. Apprentices are seated and interact within the virtual environment through a head mounted display (HMD) and cabled joystick (see Figure 1). It takes approximately two and one half hours to complete VEST. Apprentices are given two attempts to achieve a minimum score of 70% to pass a performance check.

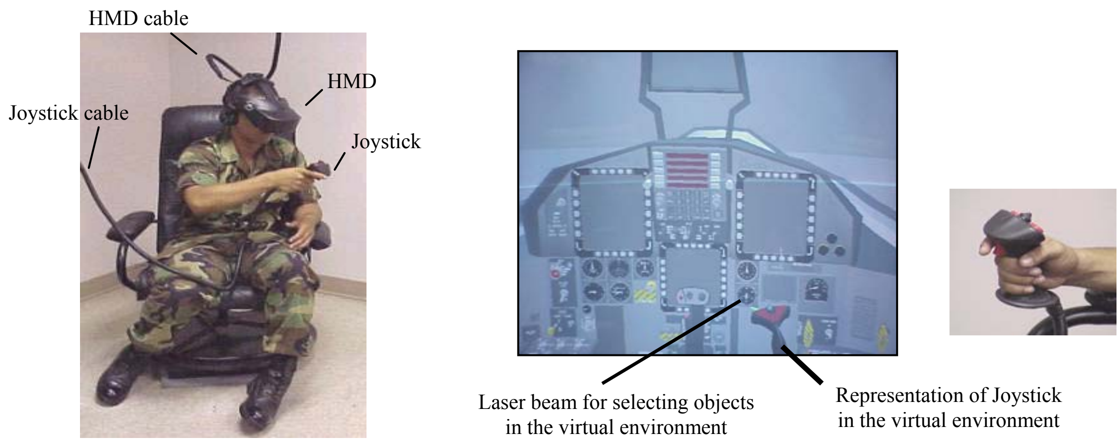


Figure 1. VEST system in operation

The Generalized Operations Simulation Environment

The Generalized Operations Simulation Environment (GOSE) project is a collaborative effort between AETC and AFRL to develop common, cost-effective, generalized VR training platforms for aircraft maintenance training, undergraduate pilot training, and space and missile training. Our initial focus is the aircraft maintenance training domain since it provided an opportunity to build on the VEST architecture. Development of GOSE involves re-engineering VEST as a scalable, modular, immersive VR training system comprised of PC-based hardware and software. GOSE initiatives include: (a) formalize training needs across airframes, (b) design instruction to meet common training objectives, (c) create a common training platform that supports multiple weapon systems, (d) increase personnel and equipment safety; (e) increase readiness in technicians trained without direct hands-on experience, and (f) reduce training costs using “virtual” replacements of aircraft assets and lower life-cycle and sustainment costs of VR platforms.

2 Evaluation of VEST

The 363rd TRS is continually assessing the current VEST system. Results from the assessment of VEST have helped us identify needed improvements to the GOSE architecture. The ongoing assessment of VEST encompasses apprentice performance scores in the VR environment and their reactions to the training experience. Apprentice technicians are asked to provide comments about their experience, strengths and weaknesses of the system, and any adverse physical effects encountered.

Performance data and comments collected from 84 apprentices assigned to complete VEST are reported here. Two apprentices were unable to complete VEST and their data was excluded from the analyses. The average performance scores across the VEST drills by number of trials is depicted in Table 1. The system failed to capture “aircraft safe-for-maintenance drill” scores for three apprentices so their overall performance scores could not be computed. The overall average performance score was 80.3%. Six of the 82 apprentices had overall average scores below 70%. Performance results indicated apprentices had the greatest difficulty with the “left console aft cockpit switch drill.” Even after the maximum number of trials (two) only seven apprentices were able to pass with acceptable scores (100%). "Review aft cockpit drill" and "weapons station drill" also indicated some degree of difficulty for 19 and 56 of the 82 apprentices, respectively. After two tries these apprentices scored only slightly over the 70% pass rate. Practice had the largest impact on the "aircraft safe-for-maintenance drill" where those who required a second try at attaining a passing score outscored those who passed on a single try ($t_{(75)} = 3.34$; pvalue = .001).

Table 1. Average performance of scores on VEST drills by number of tries (n=number of apprentices)

	Number of Tries			Number of Tries	
Drill	**1**	**2**	**Drill**	**1**	**2**
Left console forward cockpit switch drill	81.6 (n=53)	81.1 (n=29)	Left console aft cockpit switch drill	–	25.6 (n=82)
Main console forward cockpit switch drill	82.0 (n=59)	82.1 (n=23)	Main console aft cockpit switch drill	91.4 (n=77)	90.0 (n=5)
Right console forward cockpit	90.0 (n=77)	82.6 (n=5)	Right console aft cockpit switch drill	84.4 (n=73)	81.2 (n=9)
Review aft cockpit	84.4 (n=63)	74.2 (n=19)			
Weapons station drill	99.9 (n=26)	77.7 (n=56)	Safety device drill	87.1 (n=73)	89.0 (n=9)
Aircraft safe-for-maintenance intro drill	91.6 (n=79)	95.5 (n=2)	Aircraft safe-for-maintenance drill	76.4 (n=22)	84.8 (n=55)

Apprentice comments were assigned a category and a valance (positive, negative, or neutral) by two raters. Fourteen coding categories were used that encompassed all 263 comments (Table 2). Cohen's index "kappa" was used to derive inter-rater reliabilities, which were .95 for category assignment and .90 for valance assignment. Raters came to an agreement on divergent category and valance assignments, after making initial ratings, before computing the number of comments within categories.

Table 2. Category Definitions

Category	Definition
Aircraft:	comparison between VEST and F-15E
Content:	reference to instructional material
Experience:	reaction to VEST
Fix:	suggestion to improve VEST
HMD:	reference to VR HMD
Learn:	testimony to learning something
Learning Style:	preference for an instructional method
Media:	reference to training tool
Physiology:	physical reaction to VEST
Pointer:	reference to interface (hardware/software) for object selection

Unreliability:	reference to system glitches
Usability:	reference to interactions with VEST instructional method
Visual:	reference to quality of visual experience/graphical representation of aircraft
Voice:	reference to quality of aural experience

Table 3 presents the frequencies of comments by category and valance. On average, apprentices made 3.1 comments (range 1-7). The average number of positive comments was 1.3 (range 0-4). The average number of negative comments was 1.4 (range 0-5). The average number of neutral comments was .4 (range 0-2). The comment distributions were all positively skewed, showing fewer people made increasing numbers of comments. Categories are divided in the table between mostly positive, mostly neutral, and mostly negative comments. Overall, 41.5% of comments were positive, 13.3% were neutral, and 45.2% were negative. The division of comment categories by valance reveals 107 positive comments distributed over 5 distinct categories and 106 negative comments distributed over 8 distinct categories. All comments (23) in the "Fix" category, except one, were assigned a neutral valance table.

Table 3. Number of Apprentice Comments by Category and Valance

Mostly Positive				**Mostly Negative**			
Category	**Positive**	**Negative**	**Neutral**	**Category**	**Positive**	**Negative**	**Neutral**
Experience	39	8	8	HMD	—	5	—
Learn	37	—	1	Learning Style	—	12	1
Media	18	2	1	Physiology	—	20	—
Aircraft	8	—	—	Pointer	—	27	—
Constant	5	3	1	Unreliability	—	9	—
COUNT	**107**	**13**	**11**	Usability	2	21	—
Mostly Neutral				Visual	—	10	—
				Voice	—	3	—
Category	**Positive**	**Negative**	**Neutral**	**COUNT**	**2**	**106**	**1**
Fix	—	1	23				

No negative comments were made about learning or the value of gaining experience with an E-model aircraft. Example positive comments are "The VR program served its purpose very well and was extremely educating" and "gives you a realistic view of the aircraft and the loading stations." No positive comments were made about the peripheral devices, graphical representations, voice synthesis, physiology response, reliability, and learning styles.

A median split was conducted to create two groups of apprentices—those who performed at and below 50% of the class (average overall score of 80.5%) and those who performed above 50% of the class. Chi-Square tests revealed no difference in the types of comments apprentices made whether they performed at or above the median score or below the median score. Nor were differences found in the number of positive and negative comments across categories based on the median split. No relationship was found between overall performance scores and number of comments made (r = .099). Apprentices were divided into three groups based on a comparison of the number of positive versus negative comments (+ > -; + = -; + < -). No significant differences were found in performance scores across the three groups [$\underline{M}_{+>-}$ = 81.3 (n = 30); $\underline{M}_{+=-}$ = 80.1 (n = 17); $\underline{M}_{+<-}$ = 79.5 (n = 32)]. (Wenzel, Castillo, & Baker 2002).

3 GOSE Development Methods and Tools

Technical specifications for VEST are presented in Table 4. A baseline configuration for VEST was completed before start of the GOSE migration. The baseline configuration included: video capture, data backup, tutorial evaluation and demonstration.

Table 4. VEST technical specifications.

	Hardware	Software
Instructor Workstation (1)	Silicon Graphics O2 Workstation 180MHZ R5000SC (secondary cache) • 128MB RAM, 4 GB Hard drive • Mouse, keyboard, monitor • 10 Base-T to BNC multi-port repeater • DAT tape SCSI Backup System • 10 DAT tapes (4mm DAT-4-8 GB)	• Network File System (NFS) • IRIX version: 6.3 operating system • C++ compiler for Irix 6.3 • VEGA™ SP Development System • Ez3D Modeler
Student Workstations (2)	Silicon Graphics Indigo2 Maximum IMPACT (Indigo2) workstation, 195HZ R10000 • 128MB RAM, 4 MB texture memory • 4 GB hard drive • 6 DOF Motion tracking device • 2 receivers with cable • 2 flock of birds electronics units • 1 flock transmitter with 10 foot cable • HMD system: Virtual Research V6 with control box • Two Flock-of-birds model 6 degrees of freedom • Cyberwand, serial version	• IRIX version: 6.2 operating system • VIVIDS 0.82b for the SGI™ • Truetalk™ Text-to-Speech engine • Vega™ 3.0 SP Runtime • Performer 2.0.4

Table 5 contains technical specifications for GOSE. GOSE is being developed as a PC-based platform enhanced with data gloves, HMD and tracking system that will provide end-users six degrees of freedom in movement. These specifications in part address issues raised in the evaluation of VEST. The enhancements are expected to improve navigation within the virtual environment (see figure 2).

Table 5. GOSE technical specifications.

	Hardware	Software
Instructor Workstation (1)	IBM Z Pro, 2 GHZ (2 processors) • 2 GB RAM, GeForce4™ video card • Mouse, keyboard, monitor	• XP operating system • Vega™ Prime
Student Workstation (1)	IBM Z Pro, 2 GHZ (2 processors) • 2 GB RAM, GeForce4™ video card • *Virtual Research V8* HMD • Two *Flock-of-birds* • Magnetic tracking system from Ascension • Pinch Gloves • Phantom™ haptic device • CrystalEyes stereographic glasses	• XP operating system • Vega™ Prime • GHOST™ software development kit

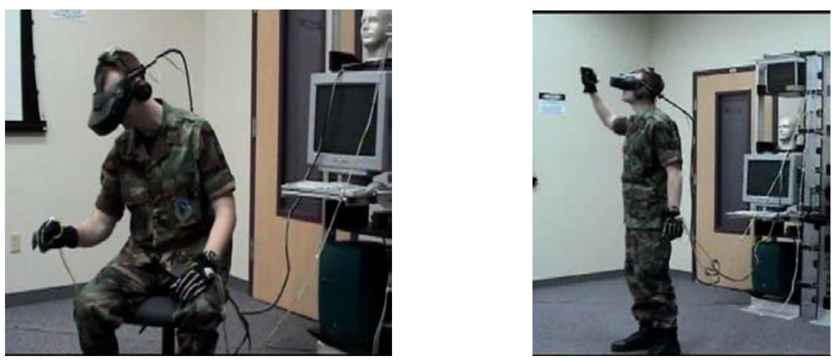


Figure 2. Concept for GOSE immersive system with improved HMD and data gloves.

Inclusion of a haptic (feedback) interface and stereoscopic glasses, along with use of three-dimensional models and synthetic voice commands, in the next phases of development should provide the end-user greater flexibility and increased realism in the immersive environment. The improvements are expected to further increase training effectiveness and reduce cybersickness. Evaluation of GOSE is planned following completion of the each phase of development. For comparative purposes, assessment criteria used to evaluate GOSE will include the same criteria used to evaluate VEST.

4 Discussion

The VEST evaluation results indicate graphical representations, joystick pointer, and cybersickness are end-user concerns. The design of GOSE is intended to address these issues and heightened usability for the end-user.

Commercial-off-the-shelf (COTS) applications to be used in GOSE accept embedded graphics (photographs), rather than constructed graphics in creating the virtual airframe. Reductions in motion lag time, making the airframe "solid," and surround detail are needed enhancements to sensory realism. It is likely that such changes in GOSE will reduce feelings of nausea and dizziness. The multimedia resource elements of VR (diagrams, images, text, video, etc.) add richness and depth to the learning experience and can be used to facilitate strong cognitive links (Hoffman & Murray, 1999). For example, training on part replacement inside an airplane fuselage or an automobile engine can be clarified by making the virtual machine's outer layers invisible or transparent.

Peripheral devices, e.g., Pinch Gloves™ are available to allow users to interact with the virtual airframe in manners similar to how technicians interact with an actual airframe. The VEST joystick will be replaced in GOSE with data gloves to enable the experience to more closely resemble hands-on training. Apprentices will be required to reach out to interact with switches and ground maintenance safety pins. Apprentices can be seated when working in the cockpit and agile when *safe-ing* the aircraft for maintenance.

In the VEST trainer, apprentices were encouraged to take breaks every 15 to 30 minutes and not to stay in longer than 45 minutes; however, some chose to complete the exercises without breaks. Reported adverse physical reactions to the VR experience such as headache and eyestrain, may be due to the length of immersion (Gupta, Klein, & Wantland, 1996). The length of immersion may also affect the instructional effectiveness of GOSE. There are reasonable points in training content for forced breaks that would shorten immersion time and support learning. In response to end-users reactions, forced-breaks in the GOSE training are: (a) cockpit switch familiarization, (b) weapons station and safety device identification, and (c) ground safe-for-maintenance procedures.

VR adds to learning through experience. End-users learn "how to" and perform physical tasks in real-time in the virtual environment without risk to an apprentice's safety (Eline, 1998). However, instructional improvements are needed in GOSE to (a) control for information overload, (b) expand content areas, (c) extend opportunities to practice, (d) guide learning, and (e) increase accessibility to the trainer. The addition of working collaboratively in the virtual environment would further enhance the learning experience and training capability of GOSE.

5 Conclusion

Results from the VEST assessment help lead the way to GOSE, beginning with addressing realism and usability issues and guiding necessary improvements to the instructional design. GOSE will provide a VR platform to continue the research to better understand the physiological responses (e.g., headache, eyestrain, disorientation, nausea, muscle stress) to the virtual environment and to continue to explore cutting-edge methods and tools to increase training effectiveness and training transfer. Assessment of GOSE should include data to determine the extent to which training with VR systems transfers to the operational environment. Elements of the operational environment such as visual and auditory representations of the flight line incorporated in the virtual environment would likely enhance transfer of training.

6 References

Eline, L. (1998). A virtual reality check for manufacturers. American Society for Training & Development: Technical Training, 9, 10-14.

Gupta, S.C., Klein, S.A. and Wantland, C.A. Cyberpathology: Medical Concerns of Virtual Reality Applications. [online]. Available: http://www.webmed.com/mi/cyberpath.html

Hoffman H., Murray, M. (1999). Anatomic VisualizeR™: Realizing the vision of a VR based learning environment. Medicine Meets Virtual Reality, 62, 134-140.

Wenzel, B., Castillo, A., & Baker, G. (2002). Assessment of the Virtual Environment Safe-for-Maintenance Trainer (VEST). United States Air Force Research Laboratory, Technical Report Number: AFRL-HE-AZ-TP-2002-0011.

Author Biography

Terence Andre is the Warfighter Skill Development & Training Research Branch Chief at the Air Force Research Laboratory, Mesa, Arizona. He directs the Distributed Mission Training (DMT) research and development (R&D) program for the Air Force Materiel Command's premier warfighter training research branch, integrating R&D efforts with users, customers, decision makers, DoD, industry, and support contractors. Lt Col Andre manages a $10M training research budget in support of MAJCOM training needs and supervises 18 government members and 35 support contractors. He received a PhD in Industrial Engineering from Virginia Tech.

Winston Bennett is a Senior Research Psychologist with the Air Force Research Laboratory Human Effectiveness Directorate, Warfighter Training Research Division, located at Williams-Gateway Airport, Mesa AZ. He is the team leader for training systems technology and performance assessment research and development. He received his Ph.D. in Industrial Organizational Psychology from Texas A&M University. Dr. Bennett has published numerous research articles, book chapters and technical reports in the Human Resources arena. He is actively involved in research related to performance evaluation, personnel assessment, training requirements identification, and quantifying the impact of organizational interventions - such as interactive, high fidelity immersive simulation environments and job redesign/restructuring and training systems - on individual, team and organizational effectiveness.

Anna Castillo is a Research Psychologist at The Air Force Research Laboratory, Warfighter Training Research Division in Mesa, Arizona. Her research goals and objectives are to utilize a DMO environment to enhance learning and performance, and determine the extent to which lessons learned in DMO transfer to the operational unit. Anna received a Master of Science in Experimental Psychology from New Mexico Highlands University in 1998 and a Bachelor of Arts in Journalism/Mass Communications from New Mexico State University in 1995.

Mary Graci is the Director of the Maricopa Institute for Virtual Reality Technologies and Adjunct Faculty at Chandler-Gilbert Community College. Mary received her Masters Degree in Education from Northern Arizona State University and a Bachelor of Fine Arts degree in Art Education from Arizona State University. She is the pioneering spirit of the virtual reality technology program, developing education and training for college credit and corporate curricula for the largest and prestigious community college district system in North America.

Dale McClain has over 6 years experience in Graphic Design, Computer Science and Computer Animation. He has been a graphic designer, artist, animator and programmer for the Air Force and Eberly College of Business. Currently he is finishing his Bachelor of Science degree in Graphic Information Technologies at Arizona State University East. He plans to continue his education with a Masters in Science.

Mathew Purtee is a Warfighter Training Research Analyst for the Air Force Research Laboratory. His key contributions involve verbal protocol analysis and integrating virtual reality with maintenance training. He has worked as a laboratory assistant for the McSweeney Laboratory and as a student therapist for the Pacific Northwest Autism Clinic at Washington State University. Mr. Purtee earned a Bachelor of Science in Psychology from Washington State University.

Brenda Wenzel was formerly Team Lead of the Collaborative Learning Instructional Technology Team at the Warfighter Skill Development & Training Research Branch, Air Force Research Laboratory, Mesa, Arizona. She is currently the Senior Engineering Psychologist for the Systems Training Division, TRADOC Analysis Center, White Sands Missile Range, New Mexico. She received her PhD in Psychology from New Mexico State University, Las Cruces, New Mexico.

Keep systematic training system design efficient![1]

Daniëlle M.L. Verstegen, M.Sc.
TNO Human Factors,
Dept. of Training and Instruction
Kampweg 5/P.O. Box 23
3769 ZG Soesterberg
The Netherlands
Verstegen@tm.tno.nl

Dr. John C.G.M. van Rooij
Royal Netherlands Army,
Staff-OTCO/PL/BO/HBBO
Mineurslaan 500
3521 AG Utrecht
The Netherlands

Summary

Systematic methods for training system design have been available for years. Using such methods for the development of user requirements and specifications for technically advanced training means (simulators, E-learning, CBT, etc.) ensures that these will fulfil the training needs. The problem is that applying systematic methods takes time and effort, and requires expertise. Our solution is to work iteratively and in workshops under supervision of experienced facilitators. The same systematic method for training system design can be used at different moments with different levels of detail. The SLIM method has been designed for the needs statement phase. In this early stage training needs analysis and training program design are done at a rather global level. Later on, using the same or a similar step-by-step method, the results can be elaborated in more detail, in order to write specifications and -eventually- to implement training. The SLIM method emphasises aspects that are often not systematically addressed: the place of training means in overall training, the choice between high and lower fidelity options, facilities for instruction, feedback and assessment, and the role of the instructors. Inviting all stakeholders ensures that the chosen solution is widely supported. In this paper we briefly describe the SLIM method, and the experiences during its application in a number of different cases. We conclude that needs statements can be developed quickly and systematically using the SLIM method in workshop setting.

[1] Parts of this paper will also appear in Verstegen (in press).

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held in Genoa, Italy, 13 – 15 October 2003, and published in RTO-MP-HFM-101.

1. Background

In the Royal Netherlands Army (RNLA) the development of a needs statement is the first phase in the procurement of technically advanced training means such as simulators, Computer Based Training (CBT) web-based learning environments and Virtual Environments. A needs statement describes a preliminary selection of training means and a first estimate of how much they will cost. Of course, it will have to be elaborated, adapted and revised later on. A well motivated needs statement is important, however, as financial, and other resources for these expensive training means have to be reserved well in advance (Verstegen, Barnard, and van Rooij, 1999). Based on discussions with stakeholders from the RNLA, the assignment for TNO Human Factors (TNO-HF) was to develop a method to be used to develop needs statements, or more specifically, to specify a way to determine the present and future training needs within a limited time span and based on information that can be incomplete or insecure, to decide whether and why the acquisition of advanced training means will be necessary, to estimate roughly the resources that should be allocated for this purpose and which other consequences (organisation, logistic and personnel) have to be taken into account.

Field research suggests that currently training needs are often not the major focus during the design of training systems, possibly because those responsible for their specification are usually operational staff or the designers of the operational systems and not the people who are responsible for training. When instructors are involved they are Subject Matter Experts (SME) who have little didactic background and experience with instructional design. The result is that specifications are often technology-pushed and product-oriented, i.e. written in terms of a simulation of the operational system or describing an existing training system seen elsewhere. In other cases, the functional specifications remain vague for a long time, and elaboration in more detail is postponed. This may cause unacceptable delays in the availability of training facilities (Farmer, Jorna, Riemersma, van Rooij and Moraal, 2000; Verstegen, Barnard and van Rooij, 1999; Wallace and Northham, 1998). Using a systematic method can help to avoid these problems.

Different kinds of systematic instructional design and development methods are described in the literature (see e.g. Verstegen, in press, for an overview). None of these methods was directly applicable, however, because they are not specifically geared towards this early phase in the acquisition process, not adapted to the context of the RNLA organisation and not integrated with other RNLA procedures. After an inventory of available method and tools, we evaluated two of them with RNLA target users: the MASTER method (Farmer et al. 1999)[2] and the BOOT decision-making support tool (van der Hulst, de Hoog & Wielemaker, 1999)[3]. The results are described in van der Hulst and Verstegen (2000). The conclusions were that the MASTER method with its many steps and sub-steps is too detailed for the development of needs statements. However, the overall structure of the method is a suitable framework. The second half of the method is specifically geared towards the specification of training simulators, but it seems that the specification of other kinds of training means can take place along the same lines, especially in this early phase. Like the MASTER method, already existing RNLA procedures, for example for the development and implementation of courses, and the design of other kinds of training means, are based on an Instructional System Design (ISD) approach (see e.g. Gagné, Briggs and Wager, 1992). Therefore, we did not foresee major integration problems. The kind of explicit decision-making support offered by BOOT was considered a valuable addition, and was initially implemented by defining explicit decision points with accompanying paper-based checklists. In future, software tools to support decision-making and/or expert systems to for specific decision points may be developed.

One of the problems of the ISD-based approach is that it takes a considerable amount of time and effort to get insight in training system specifications. If it is used too early in the acquisition trajectory the invested

[2] The MASTER method for the specification of training simulators was developed within the European defense research project MASTER (Military Applications of Simulator and Training concepts based on Empirical Research, EUCLID, RTP 11.1).
[3] BOOT stands for decision support for the selection of training means (in Dutch).

time and effort may be wasted, e.g. because the project does not get funding or because conditions change. Waiting too long, however, means that there is not enough time left for thorough analysis of training needs and alternative training solutions. Our solution is to work iteratively, at different levels of detail. In the early stages, training needs analysis training program design and training system specification are done at a rather global level, sufficient to write a well motivated needs statement. Later on, going through the same or a similar sequence of steps, the results can be further elaborated.

Another problem with ISD-based methods is that they are described in a rigid, linear way and need to be adapted to the design project context. Descriptive research indicates that this is not necessarily a problem for professional instructional designers, but is difficult for novices. Moreover, the development of needs statements for advanced training means is a long, complex process involving several parties with different backgrounds and (sometimes conflicting) interests and different kinds of (often contradicting) constraints. Our solution for these problems is to invite all important stakeholders to workshops, which take place under supervision of an experienced instructional designer who organises and monitors the design process, applying the systematic method described Section 2. The participants bring in their own information and expertise and they eventually take the decisions collectively, thus ensuring that the designed solution will be acceptable for all stakeholders.

In Section 2 we will describe our method for the development of needs statements. The method has been applied in five different cases. We will reflect on our experiences during these projects in Section 3, and present some conclusions in Section 4.

2. The SLIM method

The SLIM (Specifying Learning means in an Iterative Manner) method is based on the MASTER method, but adapted for the needs statement phase and the first phase in the acquisition process. The terminology and names of steps have been customised for the RNLA. The SLIM method emphasises aspects that are currently often not systematically addressed: the place of training means in overall training, the choice between high and lower fidelity options, facilities for instruction, feedback and assessment, and the role of the instructors.

2.1 The steps of the SLIM method

The SLIM method leads users through a global analysis and design process step-by-step. This does not, however, mean that the process will always be the same. Depending on, for example, the complexity of the domain and the amount of information available, steps may take more or less time, some steps may sometimes get less emphasis or may only be partly executed. The method consists of four main phases, all divided in sub-steps (see Figure 1):

I. Problem definition: inventory of the intended outcome/results, the demands, available resources and project risks.
II. Analysis of training needs: global analysis of needs for education and training.
III. Design blue print & Selection of training means: design first draft of training and select a suitable combination of training means.
IV. Design of specifications & Cost estimation: definition of global specifications and estimation of the resources that will be required for the acquisition and use of the training means.

Developing a complete and thorough needs statement takes time and manpower. Therefore, the SLIM method starts with an inventory of the situation in order to decide whether it is worthwhile to invest in such an undertaking. This phase is meant as an explicit analysis of the problem situation, as has been observed in the behaviour of expert instructional designers (e.g. Rowland, 1992, 1991; Goel and Pirolli, 1992, 1989). Expert designers have been observed to generate several potential solutions, and not commit to one particular solution until later on. This kind of behaviour is stimulated by asking users to sketch a number of likely training solutions in general terms up front, during the first phase (see also the discussion below). In the second phase a global analysis of the needs for education and training is executed, and in the third phase a first draft of the course or training program is designed. For the part that will be executed with advanced training means, prototypical lessons/scenarios and types of instruction and feedback are specified. This information is used in the fourth phase to determine global specifications for the required training means (primary user requirements) and other consequences for the training environment, such as the required number and capabilities of instructors, and the additional hardware, software and technical support required for training. This results in a rough estimate of the required resources. Three kinds of resources play a role here: resources for the acquisition of training means, resources for the use and maintenance of those products in future training, and resources for the development of (detailed) specifications. Figure 1 provides an overview of the four phases of the SLIM method and their steps.

The intention of the SLIM method is not to develop final and fine grained specifications, but to start thinking about how education and training will probably take place and which kind of training means will probably be used. On this basis a reasonably realistic cost estimation can be made, taking several alternatives into account. In principle, the same steps of the SLIM method can be executed at a more detailed level later on in the acquisition trajectory. We will come back to this issue in Section 2.4.

2.2 Specific aspects of the SLIM method

The SLIM method proposes solutions for some of the problems with the MASTER method encountered during previous evaluation studies (Verstegen, in press). We will discuss some of these issues below:

a) Dealing with resources
The MASTER method is not entirely clear about when and how designers should deal with information about the resources available for the acquisition and use of training means. The SLIM method starts with an explicit inventory of demands, resources and risks (Step I-2). These are stored in three lists are kept available at all times. After each step users are reminded explicitly to keep these lists up to date.

b) Generating alternatives
As discussed above, users are asked to sketch a number of possible solutions in general terms up front. This is part of Step I-1 and is meant both to make the participants' preconceived ideas explicit and as an inventory of alternative solutions. Comparing alternative solutions comes back as an explicit step half way through the design phase (see below, under point d). For now, it is up to the workshop facilitators to make sure that the participants keep on considering feasible alternatives during the rest of the design process (see Section 4).

c) Explicit decision points
After each step an explicit decision point is introduced. A general checklist with recurring questions or considerations has been developed (e.g. "Are your assumptions still correct?"), and at important decision points more specific questions are asked (e.g. after the selection of training means users are prompted to consider possible alternative solutions).

d) Prioritising and weighing alternatives
To encourage users to consider alternatives they are explicitly asked to select a wide range of suitable training means for each learning activity, and to prioritise these options from most desirable to least desirable (this is part of Step III-2). Subsequently, users select a combination of training means with which all learning activities can be executed, though maybe not all in the most preferred way. Users are advised to envision alternative combinations and to define their advantages and disadvantages, before making a choice. Later on in the acquisition trajectory these alternatives can be reconsidered, for example when the resources do not permit the acquisition of (a part of) the initially chosen set of training means.

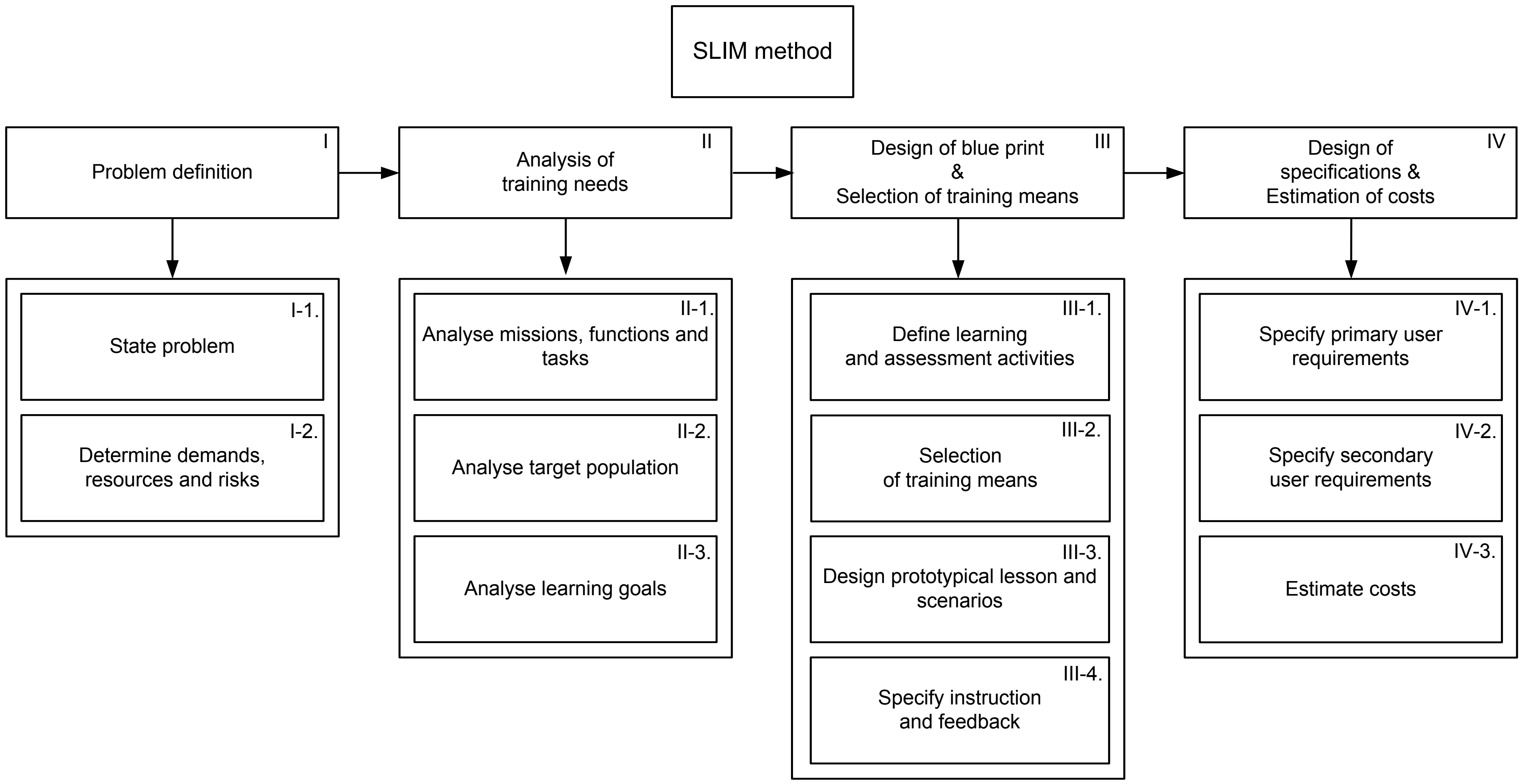


Figure 1: The steps of the SLIM method for the development of needs statements for technically advanced training means.

e) Assessment of the trainees' performance
Evaluations with the MASTER method indicated that assessment planning without an overview of the whole course is difficult, and that there is some overlap between assessment and debriefings (Verstegen, in press). In the SLIM method another point of view has been chosen: assessments are regarded as a special category of learning activity and are planned in the same step. Subsequently, they are elaborated in the same way, i.e. a range of suitable training means is chosen and, if the assessment activity is allocated to advanced training means, a 'scenario' for the assessment is described in step III-3.

2.3 Applying the SLIM method in workshop setting

The SLIM method has been applied in the form of workshops with all stakeholders under supervision of two experienced facilitators. In this setting, the workshop leader is fully responsible for the management of the design process: explaining the goals of the SLIM method, organising the process, structuring and guiding the discussions, making sure that no available information is neglected and that all feasible alternatives were taken into account, deciding when to go to the next step and also when iteration is necessary. A second facilitator takes notes and is responsible for the documentation of the entire process.

The role of the workshop participants is to provide the necessary problem and domain information, and to make decisions. Ideally, the number of participants is between four and eight, including at least a domain expert, a course designer, a future instructor and a representative from the school's management. It is also possible to invite other stakeholders, such as a representative of the operational unit(s) where trainees will work in future, a CBT developer, or financial and technical experts. Too many participants is expected to make the workshops inefficient. On the other hand, inviting all stakeholders ensures a widely accepted solution.

Iteration is encouraged by organising a sequence of workshops. In between workshops the participants can reflect on the results so far, collect additional information and discuss with their colleagues. New information or discussion points are input for the next workshop, and thus taken into account immediately.

2.4 Reusing the results of the SLIM method

The SLIM method is meant to be used at a global level of detail. Initially training needs analysis, training program design, and training system specification are executed at a rather global level, sufficient to write a well motivated needs statement. The description of steps, the guidelines, checklists and other support are focussed on helping users to go through the steps in a limited amount of time (allowing iterations, see Section 2.3). At the same time it ensures that the information collected during the first phase of the acquisition process can be reused later on when specifications are developed at a more detailed level as depicted in Figure 2.

In principle, the same steps of the SLIM method can be executed at a more detailed level later on in the acquisition trajectory, e.g. to develop functional specifications, technical specifications and, eventually, to develop and implement training. Alternatively users can choose for a more specific and detailed ISD-based method, such as the MASTER method for developing simulator specifications.

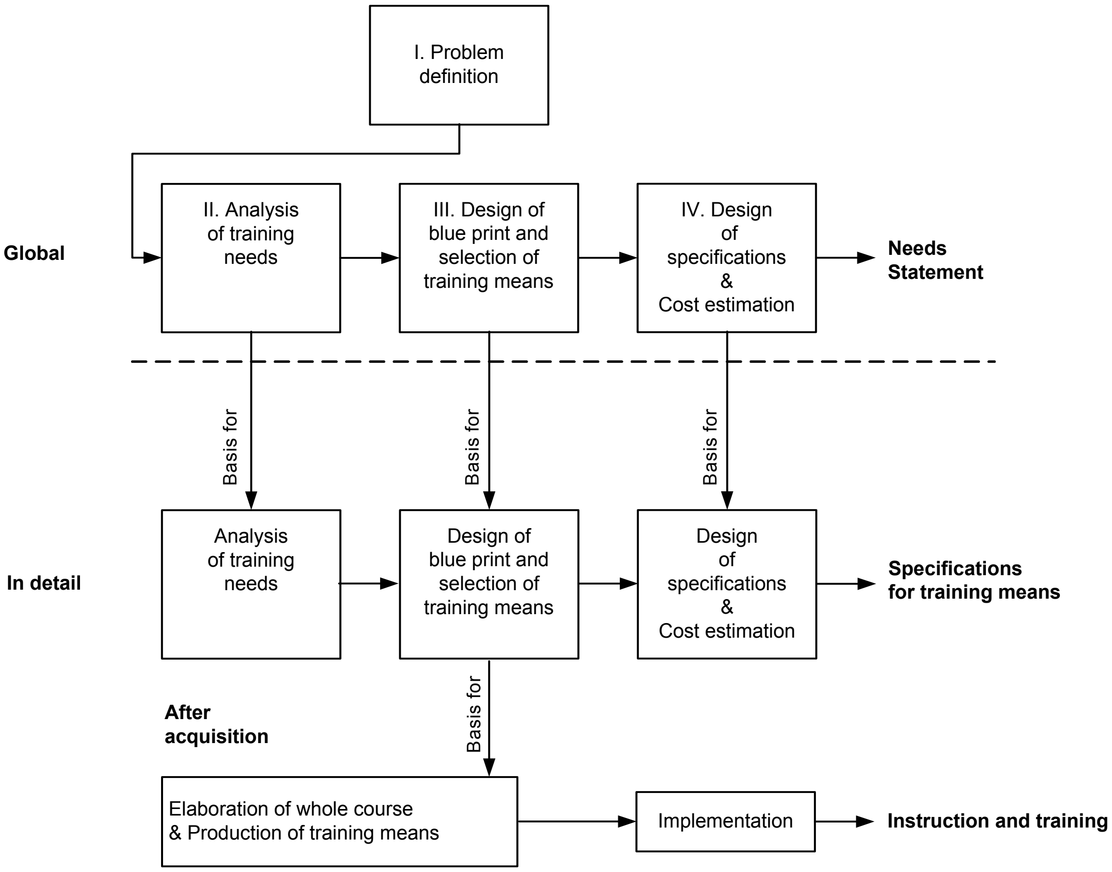


Figure 2: Reuse of information collected with the SLIM method.

3. Experiences with the SLIM method

In this section we will report our experiences with the SLIM method. After discussing the five cases in Section 3.1 we will discuss how the SLIM method was applied and adapted in Section 3.2, reflect on the role of the facilitators in Section 3.3, and briefly address iteration and dealing with uncertainty in Section 3.4.

3.1. The five cases

We have applied the SLIM method in five different cases with customers, three times for the RNLA, once for the Royal Netherlands Navy and once for the Royal Netherlands Air Force. For the last two cases, we adapted the terminology and the names of the steps to those used in these organisations. The results presented below are based on the written reports that were made of all the sessions (Boot, Verstegen and Veerman, 2002; Melis and van Berlo, 2003; Verstegen, Veerman and van der Arend, 2001), the discussions with the participants during the sessions, and the experiences of the facilitators.

In three of the cases, the resulting needs statement describes a CBT application, in one case a desktop-based simulation and in one case a simulator (all to be used in combination with other training means). All five cases were complex, but for different reasons. In one case the main source of complexity was the fact that the operational system did not yet exist, and that the information about the exact nature of the tasks that trainees would have to be trained for was incomplete and insecure. In another case the tasks were relatively easy, but the complexity from an educational point of view was caused by the very large number of trainees to be trained for different subsets of tasks. In two cases, the same training means would have to be used by clients from two rather different organisations who were planning to use the same operational system in different ways; there were also some differences in organisational culture, capabilities of instructors and the characteristics of target trainees. In both these cases, an additional complication was that the assignment of the participants of the workshop was to design CBT to be used in combination with a simulator that was being specified at the same moment by a different group (with some overlap in participants). And in one case, the course addressed during the workshop was meant to be just an example to investigate the potential role of simulation-based CBT. The challenge in this case was to incorporate an existing instructional design model for maintenance training and to replace (part of) the practice with simulators and/or operational systems with CBT-based exercises.

Since the complexity of the cases differed, the amount of time required to execute the SLIM method differed as well. The case studies took four to seven days divided over two to four sessions. In between sessions the participants went back to their workplace to collect more information and discuss the results with their colleagues. In some cases, the workshop participants took away 'homework', i.e. further elaboration of steps to be done independently in between sessions. In other cases participants were not able and/or willing to do this. It should be noted that the level of detail of the results differed as well, mainly because the amount of time available for workshops was limited by project boundaries.

3.2. The SLIM method applied to different cases

We had expected to have to adapt terminology for non-RNLA participants (see above). However, during the workshops it became clear that in all cases participants misinterpreted some terms, or used them in slightly different ways. When miscommunications hindered the process, terms were redefined on the fly or replaced by terms that were better known by the participants. In some cases the method was slightly customised as well, e.g. combining steps three and four in the design phase in a less complex case, or leaving out the fourth phase because it was not the responsibility of the workshop's participants. This ad-hoc customisation was experienced as an advantage of applying the SLIM method in a workshop setting in direct contact with the stakeholders, but it requires considerable effort and skill from the facilitators (see Section 3.3).

Applied in this way, the SLIM method proved to be robust and versatile enough for all five cases. In the experience of the facilitators the method was a valuable instrument in structuring the development process and maintaining focus and consistency. The participants valued the structured approach of the SLIM method, but clearly preferred not to be bothered with details regarding the SLIM method or instructional design and development theory in general.

Note that in this format, the participants take all the final decisions. This means that the SLIM method works towards a solution that suits the organisation and is accepted by all stakeholders, which can be –but is not necessarily– innovative. The acquisition of training means had to solve existing or expected problems of the organisation. Sometimes new types of training means were required, e.g. because the participants expected that they would have much less opportunity to let trainees practise with the operational system in future. In one case, innovation was explicitly the goal of the workshops (i.e. simulation-based CBT, see above). In another case, however, the participants decided that for a new course classroom based lessons were most suitable, much like the classroom based lessons in existing courses at the school. Another observation was that the participation of all stakeholders is vital. In one of the cases not all stakeholders were present during the workshops, which lead to uncertainty and speculation about 'what they would want', and extra time investment in reporting the results of the workshops, getting feedback and adapting the results accordingly.

3.3. The role of the facilitators

The role of the facilitators is important. The workshop leader needs to be familiar with instructional design and development models in general and with the details of the SLIM method, since he or she has to take care of organising and managing the process. The interaction can be quite unorganised, sometimes almost chaotic, since the workshop participants do not (and are not meant to) pay attention to the SLIM method and the development process. It is the task of the workshop leader to introduce the different activities (i.e. steps of the SLIM method) and to help the participants to complete the different tasks. The workshop leader also has to maintain an overview of what needs to be done, and decide when to go to the next step or back to a previous step. On top of that, the workshop leader decides if the terminology and the SLIM method need be customised, and how this can be done without violating the important principles underlying the method. The second facilitator, who is responsible for the report, also needs to be familiar with the SLIM method in order to be able to summarise and reorganise data that may result from different moments in time according to the steps of the method.

In most cases there were participants with different backgrounds and conflicting interests, and some overpowering personalities. Participants have different positions in the organisation's hierarchy, and often existing political problems played a role. The workshop format proved to be an efficient way to collect all available information, make the different opinions and arguments of stakeholders explicit, and come to joint decisions. The workshop leader, however, needs additional skills to manage these group dynamics.

It proved to be important to state the roles of the facilitators and the workshop participants clearly. In principle, the facilitators guide the process and the participants take the decisions. Only on request did the facilitators bring in their own knowledge and experience in the field of instructional design and development. Sometimes, however, giving examples proved to be a good way to get a discussion going or to stimulate the participants to consider alternative solutions.

In three cases, the participants had very high expectations of the knowledge available in the area of instructional design, illustrated by remarks such as: "So, just tell us which learning goals need CBT", "I'm sure there has been done a lot of research about how to teach these tasks", and "There must be ready-made solutions described in literature". The same phenomenon was observed in other settings, for example during evaluations of the MASTER method with target users (see e.g. van Rooij et al., 1998; Verstegen and van der Hulst, 2000). However, like other instructional design tasks, the development of a needs statement is an ill-structured problem for which there is, usually, no guaranteed best solution or solution procedure. Which combination of training means or instructional products is suitable depends not only on educational factors, but also on pragmatic factors such as the available budget, the amount and capabilities of the instructors, etc. There are no ready-made solutions available that take all these factors into account. This was an eye-opener for many participants.

3.4. Iteration and dealing with uncertainty

In all cases there was iteration within the period dedicated to the development of the needs statement, usually caused by the fact that the required input for a later step was incomplete or not elaborated into enough detail or because of new information from sources or people not present at the workshop. Sometimes, participants phoned colleagues during the workshop for missing input information. In one case, the participants proposed to split of the first phase (Problem Statement) into a one-day workshop, followed by an interval to collect additional information. The participants were less inclined to iterate in order to improve their design or to consider alternative solutions.

Uncertainty played a role in all cases, but for different reasons. In the case where the operational system did not yet exist there was much uncertainty about the tasks, especially the division of tasks between team members and the learning difficulty of the tasks. In another case the source of uncertainty was an ongoing reorganisation with -as yet- unknown consequences for the division of tasks over personnel. In a third case, the participants' task was to develop a needs statement for CBT to be used in addition to a simulator for which the specifications were being developed simultaneously by other people. However, it was not yet clear what the capabilities of the simulator would be, and indeed not even whether the simulator would be bought or not. In the fourth case, the main goal of the workshops was to develop guidelines for the application of CBT in the school. During the workshop an example course was used, and it was not clear whether CBT would be necessary or (financially) feasible for this course. And in the fifth case, most uncertainty was caused by the fact that not all stakeholders were present during the workshop.

4. Conclusions

With the SLIM method used in a workshop setting a needs statement can be developed quickly and systematically. With experienced facilitators the SLIM method is robust and applicable to different kinds of cases in different settings. During the workshops the different aspects of the design task were split up: the workshop leader organised and managed the development process, and guided the development activities and the communication between the participants. A second facilitator was responsible for the documentation process and products. Thus, the participants could focus on bringing in their own information and expertise, and on making decisions. This proved to be effective form of process management, but requires skill and expertise of the two facilitators. Inviting all stakeholders ensures that all arguments will be taken into account, and that the selected solution will be accepted within the organisation. In fact, this format can be seen as a combination of a systematic and a relational or communicative approach, as advised by Kessels (1999, 1993), amongst others.

Working with a group, rather than individually, seemed to make the process more complicated: there were almost continuous 'interventions' caused by discussions between workshop participants or by new information brought in by one of them. In most cases the most important stakeholders were present, though iteration was also caused by information or opinions from other people collected between sessions. The SLIM method provides a framework, that ensures that all important issues are systematically addressed. It also helps to manage the iterative design process. When participants bring in new information, the workshop leader decides to discuss it immediately or to come back to it at a later moment. The second facilitator makes sure that it is documented with the right step in the report. Working with a group also introduces a social aspect to the development process: during discussions not only rational arguments played a role, but also personal interests, personalities, differences in rank or power, existing conflicts or power struggles (between individual people or between different parts of the organisation), differences in interaction styles, and so forth. The workshop leader needs additional skills to manage these group dynamics. It is possible that GroupWare systems could support this aspect of the development process during the workshops.

References

Boot, E.W., Verstegen, D.M.L, & Veerman, A.L. (2002). *Behoeftestelling voor GOLMen ten behoeve van SQUIRE* [Needs assessment for advanced training means for the SQUIRE radar system] (Report no. TM-02-A029). Soesterberg, The Netherlands: TNO Human Factors (in Dutch).

Farmer, E.W., Jorna, P.G.A.M., Riemersma, J.B.J., van Rooij, J.C.G.M., & Moraal, J. (1999). *Handbook of Simulator-based Training.* London, UK: Ashgate.

Gagné, R.M., Briggs, L.J., & Wager, W.W. (1992). *Principles of Instructional Design (4th ed.).* Forth Worth (etc.): Harcourt Brace Jovanovich College.

Goel, V., & Pirolli, P. (1992). The Structure of Design Problem Spaces. *Cognitive Science 16*(3): 395-429.

Goel, V., & Pirolli, P. (1989). Motivating the Notion of Generic Design within Information-Processing Theory: The Design Problem Space. *AI Magazine 10*(1): 19-36.

Hulst, A. van der, Hoog, R. de & Wielemaker, J. (1999). *BOOT: Decision support for the selection of facilities for education and training.* (Report no. FEL-99-A188). The Hague, The Netherlands: TNO Physics and Electronic Laboratory.

Hulst, A.H., van der & Verstegen, D.M.L. (2000). GOLM ontwikkeling KL: Vormen van ondersteuning voor behoeftestelling [Specification of advanced training means within the Dutch Army: forms of support for the user requirements specification] (Report no. FEL-00-A191). The Hague, The Netherlands: TNO Physics and Electronic Laboratory (in Dutch).

Kessels, J.W.M. (1999). A relational approach to curriculum design. In: J. Van den Akker, R. Branch, K.L. Gustafson, N. Nieveen, & T. Plomp (Eds.). *Design approaches and tools in education and training.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Kessels, J. W. M. (1993). *Towards design standards for curriculum consistency in corporate education*. Unpublished doctoral dissertation, University of Twente, Enschede, The Netherlands.

Melis, P. & Berlo, M.P.W. van (2003). *Richtlijnen voor het opstellen van functionele specificaties voor computer-based maintenance training* [Guidelines for the formulation of functional specifications for computer-based maintenance training]. (Report No. TM-03). Soesterberg, The Netherlands: TNO Human Factors (in Dutch).

Merrill, M. D. (2001). Components of instruction toward a theoretical tool for instructional design. *Instructional Science*, *29*, 291-310.

Rooij, J.C.G.M. van, Barnard, Y.F., Verstegen, D.M.L., Bermejo Muñoz, J., Retamero Merino, S., Krawies, J. Hardinge, N., & Molloy, J. (1998). *Prototype-Validation Studies.* EUCLID RTP 11.1, MASTER Deliverable A2.4. Prepared for the Directorate of Materiel of the Royal Netherlands Army under contract no. DMKL/EUCLID/RTP 11.1 016-92-7211.11.

Rowland, G. (1992). What do instructional designers actually do? An initial investigation of expert practice. *Performance Improvement Quarterly*, *5*(2), 65-86.

Rowland, G. (1991). *Problem Solving in Instructional Design.* Unpublished dissertation. Indiana: Indiana University.

Verstegen, D.M.L. (in press). *Iteration in instructional design: an empirical study on specification of training simulators* [working title]. Unpublished dissertation. Utrecht, The Netherlands, Utrecht University.

Verstegen, D.M.L, & Barnard, Y.F. (2000). *Beproeving MASTER-methode in het kader van de klankbordgroep GOLM-ontwikkeling* [Evaluation of the MASTER method for the GOLM project],(Report No. TM-00-M024). Soesterberg, The Netherlands: TNO Human Factors (in Dutch).

Verstegen, D.M.L., Barnard, Y.F. & Rooij, J.C.G.M. van (1999). *De specificatie van Geavanceerde Onderwijsleermiddelen* [The specification of advanced training means] (Report no. TM-99-A044). Soesterberg, The Netherlands: TNO Human Factors (in Dutch).

Verstegen, D.M.L., & Hulst, A.H. van der (2000). Standardized development of a needs statement for advanced training means. *Proceedings of the 22nd Interservice/Industry Training, Simulation and Education Conference, November 27-30* (pp. 1136-1144). Arlington, Virginia: National Defense Industrial Association (NDIA), National Training Systems Association, (NTSA).

Verstegen, D.M.L., Veerman, A.L. & Arend, J.G.M. van der (2001). *Methode voor de behoeftestelling voor geavanceerde onderwijsleermiddelen: Evaluatie en verbetering* [Method for the development of needs statements for advanced training systems: Evaluation and improvement] (Report no. TM-01-A073). Soesterberg, The Netherlands: TNO Human Factors (in Dutch).

Wallace, Ph. & Northham, G. (1998, Issue 4). A training task analysis. *MS&T Magazine*, 1998 (4), 10-19.

Overview:
The U.S. Office of Naval Research Training Technology R&D

Dr. Susan F. Chipman
Office of Naval Research, Code 342
800 N. Quincy Street
Arlington, Virginia 22217-5660
United States of America

Summary

The U.S. Office of Naval Research (ONR) has a long, continuous, and distinguished tradition of support for research and development of advanced technologies for application to military training. This paper provides an overview of the history and structure of those research programs with emphasis on areas of major current activity: 1) basic research on tutorial dialog aiming for true natural language interaction capability for artificially intelligent training systems; 2) effective instructional strategies for artificially intelligent coaching in dynamically evolving situations such as instrument flying, shipboard command information center tasks, and shipboard damage control and 3) the integration of artificially intelligent coaches into virtual reality simulators. The content of this paper is very closely related to material available on an ONR web site designed for do-it-yourself briefings : (http://www.onr.navy.mil/sci_tech/personnel/cnb_sci/342/default.htm).

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held in Genoa, Italy, 13 – 15 October 2003, and published in RTO-MP-HFM-101.

This NATO Symposium on Advanced Technologies for Military provided a very appropriate opportunity to review the important contributions to advanced training technologies that the U.S. Office of Naval Research (ONR) has made over the years, since it was founded in 1946 to maintain the relationships with university researchers that had proved very useful during World War II. Few people recognize how important research and development support from ONR, as well as other U.S. Department of Defense research agencies, has been to the development of education and training technologies in the United States. Research in psychology, including research directed at training applications, was included in the research programs from the outset, so ONR has had a long, continuous, and distinguished tradition of support for advanced training technologies. Consequently the subject of this paper is a very large one that could easily fill an entire book. Here it is possible to provide only a rather brief overview. As of this writing, however, more information can be found on a public ONR web site that it designed for do-it-yourself briefings adapted to the interests of the user. Its current url is:
(http://www.onr.navy.mil/sci_tech/personnel/cnb_sci/342/default.htm).

ONR's Strong History of Contributions to Training Technology

In 1993, I was the invited speaker at the last meeting of the Military Special Interest Group of the American Educational Research Association, asked to speak on the contributions of DoD to education and training technology. At that time, I reviewed the complete collection of annual project books, describing the research projects that ONR had supported. (Unfortunately these books are no longer being produced.) These were quite revealing. In the late 1950's and the 1960's, ONR supported all the well-known pioneers of computer-assisted or computer-based training as we know it today. These efforts focused on the goal of individualized instruction, including mastery testing, branching instruction based on testing outcomes, and selective remedial instruction. Even today, this sort of individualization is not as frequently implemented in computer-based instruction as we would hope. In the 1970's investigations of what we could call interactive video began. The first investigations predated the invention of the interactive videodisk , using picture presentation devices designed for psychological experiments in animal learning. An early study demonstrated that pictures simulating interaction with a piece of equipment could be as effective as interaction with the real equipment in maintenance training applications. This was the beginning of a line of work described by Towne (1987). Support of research on issues of effective multimedia design, initially in the videodisk environment, began about 1980, notably the research of Patricia Baggett. Her research is summarized in Wetzel, Radtke, and Stern (1994), *The Instructional Effectiveness of Video Media,* a broad review that is itself a product of Navy training research investments.

A major emphasis of ONR's research programs over the years has been artificially intelligent tutoring, also known as intelligent computer-assisted instruction (ICAI). The first contract in this area was awarded to the late Jaime Carbonell of Bolt, Beranek and Newman, Inc., in 1969. At that time, a $2M computer was required to do the research. In the 1980's intelligent tutoring became an important program emphasis with a substantial increase in dedicated funding. During part of this period, both the U.S. Army Research Institute (ARI) and the Defense Advanced Research Projects Agency (DARPA) participated in jointly funded projects with ONR. Under the Defense University Research Instrumentation Program, ONR also provided many members of the university research community AI workstations to be used in intelligent tutoring research. By that time, the cost of the necessary computers had come down to $20-$60K each. The important role of ONR and other DoD agencies in the development of artificially intelligent tutoring technologies was very much related to the willingness to invest in long term research using computers that were, at the time, too expensive for practical instructional use. In contrast, a joint program of the U.S. National Institute of Education and the National Science Foundation, Science Education Directorate, supported R&D exploring the use of computers in mathematics education at about the same time. That program required that the researchers use computers that could be expected to be in classrooms within five years, and most of the projects used Apple II computers that actually were in classrooms at the time. Such computers were not adequate to implement artificially intelligent tutoring approaches.

By the time the special funding emphasis on artificially intelligent tutoring concluded, one could fairly say that the American, and international, research community basically knew how to build highly effective tutoring systems. Among the papers and books summarizing the accomplishments of this era are Anderson, Boyle, & Reiser (1985), Clancey (1986), Lesgold, Chipman, Brown, & Soloway, (1990), Wenger, (1987).

There were also two notable projects demonstrating that the technology could apply to practical military maintenance training problems. One of these was the Navy's Intelligent Maintenance Training System (IMTS) (Towne, 1987) that also led to research on authoring tools to build similar systems at lower costs. The other was the Air Force's SHERLOCK (Gott, 1989), which involved one of the best evaluations of the effectiveness of intelligent tutoring that has yet been done. Although SHERLOCK was an Air Force project, it built upon earlier ONR sponsored work both at the University of Pittsburgh and Bolt, Beranek and Newman. In addition, ONR funded a demonstration application to troubleshooting ship steam power plants (Vasandani, 1991). The National Science Foundation also began to support applications of artificially intelligent tutoring to education, mainly supporting researchers whose earlier work had been supported by ONR (Koedinger, Anderson, Hadley, & Mark (1995), Schofield, (1995).

In the 1990's, ONR, along with many other agencies, lept on the virtual reality band wagon. Substantial investments were made. Some relatively basic research was supported, investigating such issues as simulator sickness in virtual reality. As the decade closed, some practical training applications for the Navy, that will be described below, were beginning to emerge. Emphasis is now shifting from trainers of vehicle operation for both the Navy and Marine Corps to applications in urban warfare for the Marines.

Investments related to artificially intelligent tutoring continue. At the basic research level, current efforts aim at achieving true natural language interaction capability for artificially intelligent tutoring systems, tutorial dialog. Cognitive modeling capability is also being expanded, notably into the psychological representation of space and spatial reasoning abilities. This will expand the range of model-based tutoring systems that can be built. Somewhat more applied research projects combine investigations of instructional strategy issues with building demonstration tutors.

The Vital Importance of DoD Investment to the Advance of Training Technology

Unfortunately, our ultimate military customers for training technology do not always realize that the technology they see industry offering them today probably would not exist if the underlying research had not been supported by the DoD agencies many years before. Many presume that the technology was actually developed by industry or that the Department of Education takes care of that sort of thing. For this reason, the following quotation from Frances Degen Horowitz, as she left her position as the American Psychological Association's Chief Science Advisor, is one of my favorites. Over the years, she had learned that, "The value of the military as an apolitical setting in which to develop state-of-the–art teaching and testing devices and materials is a factor that ought to be taken into account as plans are developed for the downsizing of the military. The Defense Department funds and applies more research in teaching, learning and testing than does any other branch of the federal government." As a developmental psychologist, Horowitz herself never received DoD funding, but she had learned of its importance over the years, especially through her administrative responsibilities. *Apolitical* is the key word in Horowitz's statement. Usually there is strong consensus about what the goals of military training courses are; the same cannot always be said for civilian education. Consequently, the military has provided a more stable environment for pursuing research on instruction. Much of what is viewed as American "educational" research has actually been supported by the U.S. Department of Defense, notably by ONR. In the United States, education is generally viewed as the responsibility of local communities or of the individual states. Federal level influence on the content of the curriculum tends to be viewed with suspicion. Even the existence of a federal Department of Education is a relatively recent and insecure development. Consequently there has not been a solid political foundation for civilian educational research and development programs in the United States. The U.S. military cannot rely upon the civilian sector as a source for education and training research knowledge.

Artificially Intelligent Tutoring Technology: A Continuing Theme in the ONR Program

The advancement of artificially intelligent tutoring technology has been a major theme of ONR's training research for many years, starting with that first award to Jaime Carbonell of Bolt, Beranek and Newman in 1969. This same underlying technology can be applied in many different delivery options that are sometimes talked about as if they were different technologies: Schoolhouse training using computers, web-based training, shipboard training, embedded training built into military systems. These may look different and may pose different system integration problems, different hardware problems. But the human aspect of the technology, the nature of human learning processes and the instructional interactions that will foster human learning remain much the same.

Why has artificially intelligent tutoring remained a goal for so long? Because individualized tutoring by human tutors seems to be stunningly effective as compared to conventional classroom instruction. Bloom's important paper (Bloom, 1984) argued that it was two standard deviations more effective. However, this could be an underestimate because his evidence included tutoring by inexpert tutors. With a few exceptions, however, individual human tutoring tends to be considered prohibitively expensive, so the pursuit of artificial substitutes becomes very attractive. By now, the enterprise has enjoyed considerable success. A number of artificially intelligent tutors have shown the ability to increase student achievement by about one standard deviation (Anderson, 1993; Koedinger et al., 1995). Although not as many practical tutors as we would like have been built and evaluated, we are beginning to see many topics under the special program for small businesses that ask for the building of artificially intelligent tutors for military use, and a few have been purchased as regular training system procurements (McCarthy, Pacheco, Banta, Wayne, & Coleman, 1994). Artificially intelligent tutoring technology is ready for prime time, and computing technology has matured enough to accommodate it at very reasonable costs. Even production costs for artificially intelligent tutoring systems seem to be comparable to costs for high quality conventional computer-based training. The primary challenge now is to persuade customers for military training systems to ask for this newer technology and to assist them in being smart buyers.

Remaining Challenges for Intelligent Tutoring Research

Although intelligent tutoring research has been quite successful, we still have quite a ways to go to match the effectiveness of human tutors. Here are some of the open research issues that we see in this area:

- True Natural Language Interaction – the special features of tutorial interaction
- Effectiveness of Alternative Instructional Strategies
- Comparative Value of Cognitive Models Differing in Detail/Development Cost
- Overcoming the Cognitive Task Analysis Bottleneck
- Added Value of Sophisticated Tutoring Strategies vs. Simple Reteaching
- Effective use of Multimedia & Virtual Reality

Tutorial Dialog

At the basic research level, the one issue we have chosen to take on at ONR, despite expert advice that it was too difficult, is developing the capacity for artificially intelligent tutoring systems to do true natural language interaction with the student or trainee. Current state-of-the-art tutoring systems are very clever at avoiding the need to do true natural language interaction. It is a salient difference between them and human tutors. Therefore, it seems a promising path for closing the effectiveness gap. Our resources are very limited, so the program goals have focused on understanding student inputs and generating appropriate explanations or other language to present to the student. Neither speech recognition nor speech generation has been included in the program. We have relied on other agencies, notably DARPA, to advance those technologies. While some other agencies have supported work on artificial dialog systems, the focus on the special characteristics of tutorial dialog has been unique to ONR.

Investments began with detailed studies of the linguistic behavior of human tutors (Fox, 1993; Graesser, Person & Magliano, 1995). Next, two computational linguists who also studied human tutorial language but then also attempted to emulate it in computational systems became the key figures in the tutorial dialog program, Martha Evens at the Illinois Institute of Technology and Johanna Moore, who was initially at the University of Pittsburgh but moved to the University of Edinburgh. Over a period of about 10 years, we have made considerable progress. Notably, Martha Evens of the Illinois Institute of Technology produced what I believe to be the first example of a tutor with true natural language interaction capability that reached a level of quality good enough to be used with real students, in this case medical students, who were not being paid to be subjects (Cho, Michael, Rovick, & Evens, 2000). These results were reported immediately after the conference's keynote speaker predicted that such accomplishments were still 10 years in the future. Evens and her associates, as well as Johanna Moore, who is now at Edinburgh, have produced many interesting results about the details of the instructional strategies and linguistic forms that human tutors use (e.g. Hume, Michael, Rovick, & Evens, 1996; Moore, Lemaire & Rosenblum (1996); Core, Moore, & Zinn, (2003)). Evens' project produced a very large number of publications (listed in Evens, 2000), but most of them are in rather obscure conference proceedings. A large number of graduate students (24) also did their dissertation work under this project. Evens and her medical colleagues Michael and Rovick are now writing a book to be published by Erlbaum Associates that will sum up the lessons learned in this research effort. The subject matter domain of the tutor is cardiac physiology; Michael and Rovick are medical school professors, expert tutors who were both the subject of study and collaborators in the effort. Many of the lessons learned about tutoring language and strategies, however, are quite general and already proving very influential. Because Moore had studied human tutoring within the context provided by the SHERLOCK troubleshooting tutor, Moore's early results were used to improve the quality of feedback to trainees in the final delivered version of SHERLOCK, SHERLOCK II. The primary focus of Moore's work is now on developing a clean modularized architecture for tutorial dialog, in which different types of required knowledge are cleanly separated (Zinn, C., Moore, J.D., and Core, M.G., (2002a&b, in press 2003). Another researcher who worked with Moore has now become an independent contributor, Barbara Di Eugenio of the University of Illinois, Chicago (Di Eugenio, 2001).

Tutorial interaction involves gestures as well as speech, and the ONR program has included a small amount of investigation of the instructional use of gestures. Herbert Clark and his students are working with the Stanford MURI project discussed below, investigating gestures. Beth Littleton of Aptima Corporation did some particularly interesting work investigating how instructors and other submariners use gestures while talking about the approach officer's task, which involves very complex spatial reasoning about the localization of another submarine. The use of gestures to express uncertainty was an interesting aspect of the project findings.

Over the last 3 years, the ONR program has been augmented by two large grants with funding from the Office of the Secretary of Defense under the Multi-Disciplinary Research Initiative (MURI) program. One of these projects, directed by Prof. Stanley Peters, a computational linguist at Stanford University, is using an intelligent tutor of shipboard damage control as a demonstration platform (Fry, Ginzton, Peters, Clark, & Pon-Barry (2001). Martha Evens is a consultant to this project. (The damage control tutor, DC-TRAIN, is itself a more applied demonstration project by David Wilkins of the University of Illinois.) Peters has integrated speech recognition technology that was supported in the past by DARPA and Festival text-to-speech technology from the University of Edinburgh with his own natural language technology. DC-TRAIN provides the natural language system with information about what the student did, what errors were made, and what should have been done. The natural language system then provides an after-action review in the form of an interactive dialog with the trainee. A transcript of a demonstration system appears as Appendix A. For this demonstration a domain specific voice was built in Festival by recording quite a number of sentences in the domain: the resulting quality of the speech is so good that only a few small glitches caused by the fact that one word ("do") that was recorded only in a sentence final position reveal that it is indeed computer generated speech. Note also that the system is capable of responding to student initiatives such as the desire not to discuss a particular issue.

The second Tutorial Dialog project, Why2000, tutors qualitative reasoning about physics problems and is a collaboration between Kurt VanLehn at the University of Pittsburgh and Art Graesser at the University of Memphis (Graesser, VanLehn, Rose, Jordan, & Harter, 2001). A number of researchers who worked with Johanna Moore when she was at Pittsburgh are working on this project, as has at least one of Evens' former students. It is exploring combinations of symbolic language processing approaches with more statistical semantic approaches using Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998). It also relates to considerable prior work on physics learning that ONR has supported at the University of Pittsburgh and VanLehn's physics tutor for use at the Naval Academy, described below. At times, you may be able to communicate with the Why2000 tutor over the Internet. It communicates in written language.

Basic Research on Computational Theories of Human Cognitive Architecture

Before we turn to discussion of more applied work in training technologies, the major emphasis of ONR's basic research in Cognitive Science should be mentioned at least briefly. It is integrative computational theories of human cognitive architecture. These theories of cognitive architecture provide theoretical foundations both for artificially intelligent tutoring systems built around detailed cognitive models of the desired knowledge and skill and for cognitive engineering of effective human computer interaction. ONR has been the most important source of support for these integrative theories and supports several of them, notably John Anderson's ACT-R (Anderson & Lebiere, 1998), which now has a substantial world-wide group of users, Carpenter and Just's CAPS (Just, Carpenter & Varma, 1999; Just & Varma, 2002), and Kieras and Meyer's EPIC (Kieras & Meyer, 1997; Meyer & Kieras, 1997 a & b, 1999). Gentner and Forbus' (Falkenhainer, Forbus, & Gentner,1989; Forbus, Gentner, & Law, 1995; Gentner, Bowdle, Wolff, & Boronat, 2001) program of research on analogical reasoning and the retrieval of analogical material, with the accompanying computational modeling, has also been supported by ONR. Allen Newell's SOAR (Newell, 1990), which has been more oriented to artificial intelligence than to psychology, has been supported primarily by DARPA, but ONR has supported quite a number of projects in which SOAR was used to model human cognition.

The most important differences among these theories probably are those which arise from the range of phenomena which they have chosen to address first. From very early on, efforts in ACT-R emphasized problem solving and the learning of problem solving of the level of complexity seen in school and college courses. This made it highly relevant to instructional applications. By now, because of the large community of researchers working in the ACT-R framework, a wide range of phenomena have been modeled in ACT-R.

A web site housed at Carnegie Mellon University provides a great deal of information about this work. EPIC set out to account for the major phenomena of so-called human performance research in which people are asked to do more than one simple task simultaneously. Further, Kieras and Meyer adopted an Occam's Razor approach to the development of their architecture: nothing would be assumed in the architecture unless it was actually required in order to account for the empirical data. Although the response delays found in dual-task experiments generally have been used to argue for the existence of central processing limitation, Kieras and Meyer were able to account for the major phenomena in that literature *without* assuming any central processing limitations. They were explained, in precise quantitative detail, by competition for where the eyes are looking and what the hands are doing. These results had major impact on the cognitive modeling research community because they demonstrated the importance of integrating models of perception and motor action with models of cognition. A graduate student quickly explored putting SOAR's "mind" into EPIC's (partial) body (Chong & Laird, 1997), and ACT-R was quickly expanded by incorporating very similar modeling of perceptual and motor systems, becoming ACT-R/PM (Byrne & Anderson, 1997). In contrast, work with CAPS has emphasized situations in which processing limitations are important and serve to explain many individual differences in performance. It is to be expected that EPIC will acquire some central processing limitations when those are required to model some task performance that the research team chooses to address. SOAR was not originally intended to be a valid model of human cognition; it was intended to be a psychologically inspired computational system for work in artificial intelligence. Consequently, it retains a number of unrealistic features as a model of human cognition.

Newer investments emphasize expansion of the range of phenomena that can be integrated into the architectural theory: human representation of space and spatial reasoning, the impact of stress and associated or similar physiological variables on cognitive performance, accounting for the brain activity associated with cognitive activity. These should expand the range of training applications for which high quality cognitive models can be built as the foundation for artificially intelligent training systems. Most of this work is being done in ACT-R because of the growing size of that research community and the relative ease with which new users can learn to work within this theoretical framework due to the training opportunities being provided by Anderson's group with ONR support.

Almost undoubtedly, the reason that military agencies (ONR, DARPA) emerged as the primary supporters of this new theoretical movement was that unified theories of cognition (and perception and motor control) are important to enable practical applications within a reasonable amount of time (say 30 years). Although there are many very solid psychological laws of small scope, it is very difficult to know what they imply about the design of a complex military display or the training of a complex and demanding military task.

Research on Instructional Strategies for Intelligent Tutoring/Demonstration Systems

There are many instructional strategy questions that could, if answered, provide substantial improvements in the effectiveness of artificially intelligent tutoring. Intelligent tutoring systems are very complex systems. Many of the instructional design decisions that must be made have no substantial research basis for guiding them. One must have an intelligent tutoring system in order to investigate the effectiveness of alternative instructional strategies within that context. But building an intelligent tutoring system of realistic scope is too large an investment to be made just for the sake of research investigations. It also remains important to build demonstration intelligent tutoring systems as a strategy to promote the wider application of intelligent tutoring. Consequently, the applied research program (6.2) includes projects which combine the investigation of some generally significant research questions with the building of an intelligent tutoring system that has some value in its own right. Notable examples include a math word problem tutor produced by Sharon Derry of the University of Wisconsin, her students and collaborators (Derry, Wortham, Webb & Jiang, 1996; Atkinson, 2003) and a coach for physics homework problems developed by Kurt VanLehn at the University of Pittsburgh, in collaboration with several physics professors at the U.S. Naval Academy. Research associated with the math word problem tutor included investigations of how human tutors of adult remedial education characterize individual differences among their students, investigation of the value of using the diagrams from Sandra Marshall's schema theory of math word problem solving (Marshall, 1995; Wortham, 1996)) (also a product of ONR--supported research), and a comparative study of several different approaches for using example problem solutions in instruction (Atkinson, Derry, Renkl & Wortham, 2001; Atkinson, Wortham, Derry, Jiang & Gance, 1998; Wortham, Webb, & Atkinson, 1997). Research with the ANDES physics tutor has also involved comparative evaluation of alternative instructional strategies and efforts to develop an effective strategy for encouraging students to do self-explanation of example problems. The high learning value of self-explanation emerged in earlier ONR-supported work on physics learning (Chi, Bassok, Lewis, Reimann & Glaser, 1989) and has now been replicated several times in a variety of subject matter domains. It is possible that much of the effectiveness of human tutors comes from their ability to support students in developing their own explanations. As is easily done in such systems, the physics tutor records every interaction with the tutor, and the collaborating physics professors have gained many new insights into student problem solving behavior by reviewing the traces of problem solving activity. Relatively late in the process, however, it was discovered that audio recording is a valuable addition for formative evaluation: it revealed student confusion over explanations that had not been detected previously. Successive versions of the physics tutor have been evaluated with classes at the Naval Academy (Albacete & VanLehn, 2000; Shelby, Schulze, Treacy, Wintersgill & VanLehn, 2001). Effect sizes as measured by course exam performance have varied, the largest being .9 standard deviation improvement, but the probable reason for falling short of some past tutors is that the physics professors have been unwilling to implement mastery learning in the tutor because Naval Academy students are so tightly scheduled. We hope that the tutor will be completed soon and made commercially available. There has been considerable interest from the physics education community in general.

The effort to introduce advanced training technology to the Naval Academy also included use of Ken Forbus' CyclePad (Forbus & Whalley, 1994) by Chi Wu, a Naval Academy professor of thermodynamics. Forbus' research on qualitative physics, which underlies the CyclePad software, has been supported for many years by ONR's artificial intelligence program. The development of CyclePad itself was supported by NSF Engineering Education. Supported by ONR, Wu has emerged as the most thorough-going user of CyclePad. He has published many papers on its educational use and is publishing textbooks (Tuttle & Wu, 2001; Wu, 2002, 2003) which reflect revision of course curricula around the use of CyclePad. CyclePad enables students to deal with much more complex and realistic problems. An unclassified version of PC-IMAT (see the paper by Wulfeck) has also been used in instruction at the Naval Academy.

In the last couple of years, a new direction for these tutoring demonstrations involves tutoring in dynamically evolving situations that pose different instructional strategy issues from those that characterize tasks such as programming, or solving math and physics problems, or troubleshooting electronic systems. How does one coach or tutor in these situations without disrupting the performance? John Anderson of Carnegie-Mellon University is working with a one-person simulation of AEGIS anti-air warfare (Sohn, Douglass, Chen, & Anderson, (submitted), and Stephanie Doane of Mississippi State University is working on turning her successful cognitive model of instrument flying (Doane & Sohn, 2000) into a tutoring system. This tutor involves tracking of eye movements, which are also modeled by the cognitive model and used to assess trainee learning.

An important question about alternative approaches to artificially intelligent tutoring is now being addressed by a contract awarded to James McCarthy of Sonalysts, Inc. under the Small Business Initiative Research (SBIR) Program. The strongest evidence for the effectiveness of intelligent tutoring systems comes from John Anderson's research group. Their tutors are based on extremely fine-grained models of student knowledge, built as ACT cognitive models. In contrast, other intelligent tutors use coarse grained student models at the level of instructional objectives. Their instructional effectiveness is unknown today. The Sonalysts project will be building two parallel tutors for training Anti-submarine/Anti-surface Tactical Air Controllers (ASTACs) , differing only in the grain size of their models. Probably even the fine-grained version will not be quite so fine-grained as the Anderson-style tutors. This issue is of some practical importance because it is obvious that building a fine-grained student model will cost more. In the Sonalysts effort, the cognitive task analysis was in conducted in such a way as to provide one data point on the cost difference for the cognitive task analysis; it was not as great as might have been expected.

Authoring Tools for Intelligent Tutoring Systems

Another significant emphasis of the applied research program has been attempts to develop authoring tools that will increase the efficiency of producing such systems and thus reduce the production cost. The Intelligent Maintenance Training System (IMTS) (Towne, 1987) provided a very early example of such tools. Once an author built a simulation of an item of equipment, IMTS could automatically generate an intelligent trouble-shooting tutor by applying a generic space-splitting strategy to the equipment, perhaps modified by available information about the likelihoods of part failure and the costs of investigating various areas of the equipment. IMTS also included tools to facilitate authoring of the simulation itself, as well as various tools for automatically generating instructional routines that were not artificially intelligent. The Air Force later supported a re-implementation of most of the IMTS tools in a PC environment running under Linux, but did not re-implement the capability to generate an intelligent trouble-shooting tutor. That was done under ONR support, with the name DIAG. The DIAG capability was later ported to a Windows environment where it can now be used in conjunction with other Windows-based tools, including graphics tools that now far exceed the capabilities of the IMTS-derived tools. Work with DIAG capabilities now continues in a project funded through the Orlando training lab, "An Integrated Environment for Technical Training and Aiding". This project is exploring widely touted but under-researched concepts for advanced forms of interactive electronic technical manuals that also provide performance aiding and even training functions. They will be working on demonstration applications for the LPD-17, a new ship designed for littoral warfare that will be relying heavily upon computer-based training. Towne has also developed an authoring tool for intelligent scenario-based training. Its demonstration application was to fighting fires in high rise buildings.

Attempts were also made to convert Anderson's laboratory tools for building intelligent tutors to tools that could be used by the wider community. The initial efforts were not successful. However, one of Anderson's former students and collaborators, Ken Koedinger, has continued to work on the "Tutor Development Kit", with support from a variety of sources, including NSF, ONR and DARPA. Increasingly successful summer workshops have been offered for the past few years under NSF support. Of course, we hope that at some future date the current basic research work on tutorial dialog will culminate in authoring tools that will facilitate building true natural language interaction for tutors in many subject matter domains.

Finally, we now have a project underway in the small business program (in this case funded by the Office of the Secretary of Defense) that is building an authoring tool for case-based instruction, a form of instruction commonly used for management education and for some forms of advanced military education or training. This was inspired by some striking findings emerging from Gentner's basic research project – identifying more effective ways to do case-based instruction (Thompson, Gentner, & Loewenstein, 2000). This project, which also tries to incorporate aspects of the way cases are used at the Harvard Business School, is being conducted by Susann Luperfoy of Stottler-Henke Associates. Another such authoring tool project is focused exclusively on the use of cases in medical instruction. The courseware produced by these tools, however, will not be artificially intelligent tutoring systems.

Cognitive Task Analysis

It is generally agreed that doing cognitive task analyses needed for intelligent tutoring systems, for other training systems for jobs and tasks with significant cognitive components, and for the design of human system interaction is a major bottleneck area, still more art than engineering, requiring highly trained cognitive psychologists or cognitive scientists. Therefore we have been interested in supporting good ideas that show promise of moving this toward a more routine and reliable activity. But good ideas seem to be scarce. The most notable effort to date resulted in the CAT and CAT-HCI tools developed by Kent Williams (Williams, 2000), now at the University of Central Florida. These aid, support and record hierarchical cognitive task analyses, and were designed to go all the way down to the fine-grained production system level of analysis such as seen in Anderson's cognitive models and tutors. CAT-HCI incorporates information from the psychological research literature about execution times for many cognitive and overt actions and can be used to evaluate human computer interaction designs by predicting task execution designs. It can also predict expected learning/training times, relying on past ONR-sponsored research by Kieras (Kieras & Bovair, 1986). However, there is no magic in these tools. The analyst must still have insight into the cognitive operations used.

Another notable product in this area resulted from a NATO Study Group on Cognitive Task Analysis chaired by Jan Maarten Schraagen of TNO in the Netherlands. The original goal of the study group was to match cognitive task analysis problems with task analysis methods. This proved unfeasible because most so-called methods address only a small aspect of any given cognitive task analysis problem. Any successful cognitive task analysis requires orchestrating quite a number of these small methods. However, ONR sponsored an international workshop to address the issue. This resulted in a published book reviewing the state of the art, *Cognitive Task Analysis* (Schraagen, Chipman, & Shalin, 2000) . This effort further inspired an ONR supported small business contract for a digital library resource on cognitive task analysis. Primarily, this was intended to assist people in performing cognitive task analyses by giving them access to past cases of successful task analyses, so that they could see how problems similar to their own had been addressed. The contract was awarded to Aptima, Inc. As of this writing the resource can be accessed at ctaresource.com. It remains to be seen whether the community will support the continuing existence of this resource.

Major Applied Demonstration Projects

Still more applied than the work discussed above, are demonstration projects (6.3) that demonstrate the feasibility of advanced training technology applied to a specific training problem with a defined military customer. These are intended not to produce a completely finished training product but to reduce risk before an actual training system procurement. Often, however, training demonstrations are closer to the final product than is the case for other areas of R&D because it is desirable to conduct experimental evaluations of training effectiveness. An experimental training system must be quite finished and reliable before such experiments can be conducted. Typically, these projects are quite expensive, much more so than other projects mentioned above.

The IMAT/DSOT (Interactive Multi-sensor Analysis Trainer/Deployable Sonar Operator Trainer) line of work discussed in the paper by Wulfeck represents one of the largest, longest-running and successful investments of this type. This work received the very first Bisson award recognizing successful transitions of R&D efforts into practical fleet use, a competition of across all kinds of Navy R&D.

Another major project which received a high level investment for 3 years was the Advanced Embedded Training (AET) project (Zachary, Cannon-Bowers, Burns, Bilazarian, & Krecker, 1998). This was the culmination of the line of research on Tactical Decision Making Under Stress (TADMUS) carried out at the Navy training lab in Orlando, Florida. This system automated the team training techniques developed for use with the AEGIS anti-air warfare team. It is the first example I know about of an intelligent tutoring system for a team, as opposed to an individual. It emphasized performance assessment by analyzing keystroke and trackball movements, eye tracking on the displays and speech recognition to determine whether the right information was being asked for or communicated at the right time. Detailed cognitive models of the team members, models actually capable of performing the tasks, were built to serve as a basis of comparison with trainee performance. These were intelligent, flexible "expert systems" that could handle many different scenarios, and the total project also included scenario authoring capability. Although there was a minimal capability for feedback during performance, the emphasis was on after-action review provided by a human instructor using information provided by the automated systems. AET was intended as a demonstration of what might be done in future ships such as DD-21 or DD-X. It remains to be seen whether this type of sophisticated training capability will be implemented in future ships. The cognitive models built in this research have also been used in some further research on training with synthetic team mates. The team training research behind the AET has also received a Bisson award for fleet transition.

The DC-TRAIN system mentioned above was another such project, with a much smaller budget. It is designed for training the damage control assistant, an officer who coordinates damage control efforts. DC-TRAIN is a complex system incorporating numerical simulations of fire, the spread of fire, flooding, etc., that are physically correct. In addition, it has an expert system (of admittedly somewhat limited scope) that knows how to do damage control. As in the AET system, the expert system's recommended actions are used as a basis for evaluating trainee performance. Instructional intervention is limited to an after-action critique of the trainee's performance. Like AET, DC-TRAIN has the capability for easy authoring of a large number of different damage scenarios. As the software system became larger and more complex (over 2 million lines of code), it developed reliability problems. However, these seem to have been overcome; nearly all were traced to minor errors in the database describing the ship that caused the numerical simulations to crash. A lesson learned was that a more qualitative and less sensitive simulation approach would probably be a better choice for a practical training system procurement. DC-TRAIN's written output to the student also is not as graceful as it could be. Nevertheless, DC-TRAIN has been well received by a working group on damage control training, which judged it promising for addressing the "command and control" dimension of damage control training. The natural language debriefing dialog demonstrated under the Stanford Tutorial Dialog project discussed above makes it significantly more attractive. Despite some difficulties along the way, DC-TRAIN should probably be regarded as a successful, risk reducing concept demonstration (Bulitko, V. V. and Wilkins, D. C., 1999; Sniezek, J. A., Wilkins, D. C., and Wadlington, P., (2002); see also http://www-kbs.ai.uiuc.edu. for many more reports.)

Virtual reality has also begun to pay off in practical demonstration projects. The first successful example was VE-SUB, a virtual reality simulator providing the opportunity to practice bringing submarines into harbor. This project has moved on from a demonstration project to the actual procurement of the training systems, with simulations of quite a number of different harbors. It is being implemented in all six U.S. submarine training centers. Advancing beyond this project was COVE (Conning Officer's Virtual Environment) which integrated the virtual environment simulator with an artificially intelligent coach of ship driving. This was evaluated as effective and is moving on to practical implementation. Additional vehicle trainers of this general type are also being developed for the Marine Corps.

ONR's training R&D is doing very well in providing effective training technology to the fleet, as the Bisson awards illustrate – only a few such awards have yet been made, and training R&D has received two of them. Many other practical training applications, such as the current large investments in computer based training and web-based training, also owe their feasibility to ONR research investments, even though that connection is often forgotten and rarely recognized.

References

Albacete, P.L., & VanLehn, K. (2000) Evaluation of the effectiveness of a cognitive tutor for fundamental physics concepts. In L.R. Gleitman & A.K. Joshi (Eds), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 25-30). Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1993) *Rules of the Mind.* Hillsdale NJ: Lawrence Erlbaum Associates. (see chapter 11).

Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum

Anderson, J. R., Boyle, D. F. and Reiser, B. J. Intelligent tutoring systems. *Science*, 1985, 288, 456-462.

Atkinson, R.K. (2003) An experimental evaluation of tutorials in problem solving (TiPS): A remedial mathematics tutor. Final report on ONR N00014-02-1-0191.

Atkinson, R., Derry, S. J., Renkl, A. & Wortham, D. (2001). Learning from examples:Instructional principles from the worked examples research. *Review of Educational Research, 70,* 181-215

Atkinson, R. K., Wortham, D. W., Derry, S. J., Jiang, N., & Gance, S. K. (1998). *Beyond static worked examples: Testing the efficacy of computer delivered dynamic examples.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Byrne, M. & Anderson, J. R. (1997). Enhancing ACT-R's perceptual-motor abilities. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, p. 880. Mahwah, NJ: Erlbaum.

Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13, 4-16.*

Bulitko, V. V. and Wilkins, D. C., (1999) "Automated Instructor Assistant for Ship Damage Control," *Proceedings of the Eleventh Conference on Innovative Applications of Artificial Intelligence*, IAAI-99, Orlando, July 18-22, 1999, 778-785.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989) Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 15, 145-182.

Cho, B.I., Michael, J., Rovick, A., & Evens, M.W. (2000) An analysis of multiple tutoring protocols. *Proceedings Intelligent Tutoring Systems: ITS 2000*. Montreal. P. 212-221.

Chong, R. S., & Laird, J. E. (1997). Identifying dual-task executive process knowledge using EPIC-Soar. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society*. 107-112. Mahwah, NJ: Lawrence Erlbaum.

Clancey, W. J. (1986) From GUIDON to NEOMYCIN and HERACLES in twenty short lessons: ONR final report 1979-1985. *AI Magazine*, 1986 (August), 40-187.

Core, M.G., Moore, J.D., Zinn, C. (2003) The Role of Initiative in Tutorial Dialogue, Eleventh Conference of the European Chapter of the Association for Computational Linguistics (EACL), Budapest, Hungary, 2003.

Derry, S. J., Wortham, D., Webb, D., & Jiang, N. (1996). *Tutorials in problem solving (TiPS): Toward a schema-based approach to instructional tool design.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Di Eugenio, B. (2001) Natural-language processing for computer-supported instruction. *ACM Intelligence*, 12, p.22-33.

Doane, S.M., & Sohn, Y.W. (2000) ADAPT: A predictive cognitive model of user visual attention and action planning. *User Modeling and User Adapted Interaction*, 10, 1-45.

Evens, M.W. (2000) *Language Understanding and Generation in Complex Tutorial Dialogues.* Final report on grant N00014-94-1-0338. Chicago: Computer Science Department, Illinois Institute of Technology.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence, 41*, 1-63.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*(2), 141-205. (Abridged version to be reprinted in T. Polk & C. M. Seifert, (Eds.), *Cognitive Modeling*. Boston: MIT Press)

Forbus, K. & Whalley, P. (1994) Using qualitative physics to build articulate software for thermodynamics education. *Proceedings of AAAI-94*, Seattle.

Fox, B. A. (1993). *The Human Tutorial Dialogue Project: Issues in the design of instructional_systems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fry, J., Ginzton, M., Peters, S., Clark, B. & Pon-Barry, H. (2001) Automated tutoring dialogues for training in shipboard damage control. In Proceedings 2nd SigDial *Workshop on Discourse and Dialogue.*

Gentner, D., Bowdle, B., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), (2001). *The analogical mind: Perspectives from cognitive science* (pp. 199-253). Cambridge, MA: MIT Press.

Gott, S. P. (1989) Apprenticeship instruction for real-world tasks: The coordination of procedures, mental models, and strategies. In: Rothkopf, E. Z. (Ed.), *Review of research in education*, 15, (pp. 97-160). Washington, DC: American Educational Research Association.

Graesser, A.C., Person, N.K., & Magliano, J.P. (1995) Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology,* 9(6),495-522.

Graesser, A.C., VanLehn, K., Rose, C., Jordan, P. & Harter, D. (2001) Intelligent tutoring systems with conversational dialogue. *AI Magazine*, 22(4), 39-51.

Hume, G., Michael, J., Rovick, A., & Evens, M. (1996) Hinting as a tactic in one-on-one tutoring. *The Journal of the Learning Sciences*, 5(1), 23-47.

Just, M. A., Carpenter, P. A., & Varma, S. (1999). Computational modeling of high-level cognition and brain function. *Human Brain Mapping. 8*, 128-136.

Just, M. A., & Varma, S. (2002). A hybrid architecture for working memory: Reply to MacDonald and Christiansen (2002). *Psychological Review*, *109*, 55-65.

Kieras, D.E., & Bovair, S. (1986) The acquisition of procedures from text : A production system analysis of transfer of training. *Journal of Memory and Language*, 25, 507-524.

Kieras, D.E. & Meyer, D. E. (1997) An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human Computer Interaction*, 12, 391-438.

Koedinger, K.R., Anderson, J.R., Hadley, W.H., & Mark, M.(1995) Intelligent tutoring goes to school in the big city. In: J. Greer (Ed) *Artificial Intelligence in Education, 1995,* p.421-428 Proceedings of AI-ED 95, Washington, DC: August 16-19.

Landauer, Thomas K.; Foltz, Peter W.; Laham, Darrell (1998) An introduction to latent semantic analysis. *Discourse Processes*, 25(2-3) 259-284

Lesgold, A., Chipman, S., Brown, J.S. & Soloway, E. (1990) Intelligent training systems, *Annual Review of Computer Science,* 4, 383-94.

Marshall, S. P. (1995). *Schemas in Problem Solving*. Cambridge, England: Cambridge Press.

McCarthy, J.E., Pacheco, S., Banta, H.G., Wayne, J.L., & Coleman, D.S. (1994, November) The Radar System Controller Intelligent Training Aid. Paper presented at the 16th Interservice/Industry Training Systems and Education Conference, Orlando, FL.

Meyer, D.E. & Kieras, D. E. (1997a) A computational theory of executive cognitive processes and multiple-task performance: Part 1 Basic Mechanisms. *Psychological Review*, 104,3-65.

Meyer, D.E. & Kieras, D. E. (1997b) A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory period phenomena. *Psychological Review*, 104, 749-791.

Meyer, D. E. & Kieras, D. E. (1999) Precis to a practical unified theory of cognition and action: Some lessons from computational modeling of human multiple-task performance. In D. Gopher & A. Koriat (Eds) *Attention and performance XVII* (pp. 15-88). Cambridge, MA: MIT Press.

Moore, J.D., Lemaire, B., & Rosenblum (1996) Discourse generation for instructional applications: Identifying and exploiting relevant prior explanations. *Journal of the Learning Scienc*es 5(1), 49-94.

Newell, A. (1990) *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Schofield, J. W. (1995) *Computers and classroom culture*. Cambridge University Press.

Schraagen, J.M., Chipman, S.F. & Shalin, V. (Eds) (2000) *Cognitive Task Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Shelby, R.N., Schulze, K.G., Treacy, D.J., Wintersgill, M.C., & VanLehn, K. (2001). The Andes Intelligent Tutor: An Evaluation. Paper presented at the Physics Education Research Conference, Rochester, NY, July 21-25.

Sohn, M.-H., Douglass, S. A., Chen, M.-C., & Anderson, J. R. (submitted). Unit task execution with a dynamic display in a complex problem-solving situation.

Sniezek, J. A., Wilkins, D. C., and Wadlington, P., (2002) Training for Crisis Decision Making: Psychological Issues and Computer-Based Solutions, *Journal of Management Information Science.* Volume 18, Number 4, 147-168.

Thompson, L., Gentner, D., & Loewenstein, J. (2000). Avoiding missed opportunities in managerial life: Analogical training more powerful than individual case training. *Organization Behavior and Human Decision Processes, 82*(1), 60-75.

Towne, D. M. (1987) The Generalized Maintenance Trainer: Evolution and Revolution. In: William B. Rouse (Ed) *Advances in Man-Machine Systems Research, Volume 3*. Greenwich, CN: JAI Press, p. 1-63.

Tuttle, K., & Wu, C. (2001) *Thermodynamics: A computer based approach*. Dubuque, Iowa: Kendall/Hunt. ISBN 0-7872-8902-7.

Vasandani, V. (1991) Intelligent Tutoring for Diagnostic Problem Solving in Complex Dynamic Systems. Atlanta, Georgia: Georgia Institute of Technology, School of Industrial and Systems Engineering, Center for Human-Machine Systems Research, TR CHMSR-91-4.

Wenger, E. (1987) *Artificial Intelligence and Tutoring Systems:. Computational and Cognitive approaches to the communication of knowledge*. Los Altos, CA: Morgan Kaufman.

Wetzel, C.D., Radtke, P.H., & Stern, H.W. (1994) *Instructional Effectiveness of Video Media*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Williams, K.E. (2000) An Automated Aid for Modeling Human-Computer Interaction. In J.M. Schraagen, S.F. Chipman, & V. Shalin (Eds) (2000) *Cognitive Task Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wortham, D. W. (1996). *Testing a learning environment based on a semantic analysis of math problems.* Master's Thesis, Department of Educational Psychology, University of Wisconsin-Madison.

Wortham, D. W., Webb, D. C., & Atkinson, R. K. (1997). *Effects of two curricula on proportional reasoning, problem solving, and strategy use*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Wu, C. (2002) *Intelligent Computer-Based Elementary Applied Thermodynamics*, Dubuque, Iowa: Kendall/Hunt, ISBN 0-7872-9388-1.

Wu, C. (2003) *Intelligent Computer Based Engineering Thermodynamics.* New York: Nova Science.ISBN 1-59033-359-4.

Zachary, W., Cannon-Bowers, J., Burns, J., Bilazarian, P., & Krecker, D. (1998) An advanced embedded training system (AETS) for tactical team training. In B.P. Goettl, H.M. Halff, C.L. Redfield, & V.J. Shute (Eds) *Intelligent Tutoring Systems: Proceedings of the 4th International Conference, ITS '98,* 544-553. Berlin: Springer Verlag.

Zachary, W., Ryder, J., Hicinbothom, J., Santarelli, T., Scolaro, J., Szczepkowski, M., and Cannon-Bowers, J. (1998) Simulating Behavior of Tactical Operators and Teams Using COGNET/GINA. In: Proceedings of the 7th Symposium on Computer-Generated Forces and Behavioral Representation. Orlando, Fl: Institute for Simulation and Training.

Zinn, C., Moore, J.D., and Core, M.G., (2002a) A 3-tier Planning Architecture for Managing Tutorial Dialogue, *Intelligent Tutoring Systems, 6th. Intl. Conference, ITS 2002*, Springer LNCS 2363, pages 574-584, Biarritz, France, June 2002.

Zinn, C., Moore, J.D., and Core, M.G., (2002b) A 3-tier Planning Architecture for Managing Tutorial Dialogue, System Demonstration, page 200, *Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002),* Edinburgh, UK.

Zinn, C., Moore, J.D., and Core, M.G, (in press, 2003) Intelligent information presentation for tutoring systems. In: O. Stock & M. Zancanaro (Eds*): Intelligent Information Presentation*, Dordrecht, NL: Kluwer Academic Publishers.

Appendix A

Transcript of a Tutorial Dialog for After-action Review in DC-TRAIN

Tutor: Hello, we are about to review your session from yesterday.
Tutor: All right, are you ready?
Student: Yes.
Tutor: In this session there were 5 primary crises and 1 secondary crisis.
Tutor: 5 of the 6 crises were resolved.
Tutor: Most of your actions were correct, however you made several mistakes.
Tutor: OK, let's discuss a few of those crises.
Tutor: The first crisis was a fire here in Crew Living space No. 1. [centers on this compartment]
Tutor: This fire was not contained early enough.
Tutor: As a result, the fire spread to the adjacent Combat system equipment room No. 2. [highlights 2nd compartment]
Tutor: OK, let's review how to handle this crisis. [removes highlight from 2nd compartment]
Tutor: What's the first thing you should do when the fire alarm sounds?
Student: Send repair two to investigate.
Tutor: Yes, that's right.
Tutor: However, you failed to do this during the session.
Tutor: This is a mistake you made more than once.
Tutor: Remember, alarms need to be verified first.
Tutor: OK, suppose that there is a fire.
Tutor: What's the next step?
Student: Set fire and smoke boundaries.
Tutor: Yes.
Tutor: That's what you did during the session.
Tutor: OK, what else needs to be done?
Student: The repair team should fight the fire.
Tutor: No.
Tutor: You missed a step.
Tutor: Remember, the crisis might interfere with other ship systems.
Tutor: Let's try again.
Tutor: What else needs to be done?
Student: I don't know.
Tutor: Well.
Tutor: You should order repair five to mechanically and electrically isolate the compartment.
Tutor: But you didn't do this during the session.
Tutor: This is a mistake you made more than once.
Tutor: Remember, mechanical and electrical isolation protects crucial ship systems.
Tutor: All right, suppose isolation has begun and boundaries have been set.
Tutor: What's the next step?
Student: Fight the fire.
Tutor: Yes, exactly.
Tutor: That's what you did.
Tutor: Suppose the fire has been overhauled.

Tutor: Let's now discuss the resulting smoke. [displays gray color in compartment]
Tutor: What should you do first?
Student: Send repair two to desmoke the compartment.
Tutor: Almost.
Tutor: You should send a different repair team.
Tutor: Which team should you send instead?
Student: Repair five.
Tutor: Yes, exactly.
Tutor: You should order repair five to desmoke the compartment.
Tutor: But you sent the wrong repair team during the session.
Tutor: Repair 2 was overtasked, so repair five should have been sent.
Tutor: OK, let's now discuss the flooding that resulted from your firefighting efforts.
Tutor: What's the first thing you should do in response to this crisis?
Student: Can we move on please?
Tutor: OK, let's talk about something else
Tutor: The next crisis was a flood here in Engine room No. 2. [displays compartment]
Tutor: Let's review how to handle this crisis.
Tutor: What's the first thing you should do when the flood alarm sounds?
Student: Send repair two to investigate.
Tutor: Almost.
Tutor: Let's review the jurisdiction of repair lockers.
Tutor: The flood alarm was here in Engine room No. 2.
Tutor: Please click on the location of the nearest repair locker in the left display above.
Tutor: Let me know when you're ready to continue.
[Student clicks on correct compartment]
Student: Please continue.
Tutor: Yes, that's right.
Tutor: Which repair locker is that?
Student: Repair five.
Tutor: Yes, that's right.
Tutor: You should order repair five to investigate the compartment.
Tutor: But you sent the wrong repair team during the session.
Tutor: All right, suppose that there is a flood.
Tutor: What should you do next?
Student: Set flood boundaries.
Tutor: You're right.
Tutor: However, you failed to do this during the session.
Tutor: This is a mistake you made more than once.
Tutor: Remember, setting flood boundaries helps keep floods from spreading.
Tutor: OK, what else needs to be done?
Student: I should isolate the compartment.
Tutor: Yes.
Tutor: However, you failed to do this during the session.
Tutor: We discussed this same mistake earlier.
Tutor: Remember, mechanical and electrical isolation protects crucial ship systems.

Tutor: OK, suppose isolation has begun.
Tutor: What's the next step?
Student: Can we move on to the next topic please?
Tutor: OK, let's talk about something else
Tutor: Would you like to discuss how you managed the firemain?
Student: Yes.
Tutor: What's the first thing you should do after General Quarters has been called?
Student: Order the DCCO to start a firepump.
Tutor: No, that's not right.
Tutor: You should have done something else first.
Tutor: You need to ask permission to do something.
Tutor: Now try again.
Tutor: What's the first thing you should do after General Quarters has been called?
Student: I don't know.
Tutor: Well.
Tutor: You should request permission from the EOOW to start a firepump.
Tutor: But you didn't do this during the session.
Tutor: Remember, you need the EOOW's permission before starting a firepump.
Student: Why?
Tutor: Well.
Tutor: This is specified in damage control doctrine.
Tutor: Would you like me to display the relevant section of doctrine?
Student: Yes.
Tutor: OK. [doctrine is displayed]
Tutor: Let me know when you're ready to continue.
Student: Please continue.
Tutor: OK, suppose you've received permission to start a firepump.
Tutor: What should you do next?
Student: Order the DCCO to start a firepump.
Tutor: Yes, that's right.
Tutor: You did this during the simulation.
Tutor: OK, let's review the most important points from this session.
Tutor: You performed well in the areas of investigating, containing, and eliminating crises.
Tutor: We discussed your errors relating to isolation, and permission to start a firepump.
Tutor: Remember, it's important to mechanically and electrically isolate a compartment in order to protect crucial ship systems.
Tutor: Also, you must request permission from the EOOW before starting a firepump.
Tutor: OK, that ends our review.
Tutor: Would you like to review this session again?
Student: Nope.
Tutor: Goodbye.

Training Issues in Current UK In-Service Military Equipment

Jo Davies and George Ward
ESE Associates Ltd
15, Jesse Close,
Yateley,
Hants
GU46 6AH
jo@ese-assoc.demon.co.uk

Summary

This paper presents the findings of a survey carried out during 2001/2002 to identify Human Factors issues and concerns associated with in-service equipment. Training was one of the topics under investigation and 15 of the projects interviewed out of a total of 24 projects highlighted specific training related concerns. The types of issues found were categorised as follows:

- Degree of trainer fidelity
- Maintainer training aids
- Trainer utility/utilisation
- Documentation support
- Training gaps/conflicts
- Scenario definitions

The paper will discuss the specific issues to determine which ones present future challenges in terms of training needs

Introduction

A 2-year study was carried out between 2000 and 2002 on behalf of QinetiQ CHS for UK MOD to investigate the status of HFI application in UK defence equipment procurement Management within the Defence Procurement Agency (DPA) and Defence Logistic Organisation (DLO) Integrated Project Teams. The overall aim of the study was to facilitate the allocation of HFI research budgets within the UK MOD Corporate Research Programme

The first year of the study investigated specifically the methods, techniques and organisational aspects associated with the planning of HFI within the IPTs. The 2nd year addressed HF issues associated with In-Service Equipment and it is the findings related to training matters of the 2nd year of the latter that are presented in this paper

The structure of the paper is as follows:

- Our approach
- Project Selection
- Taxonomy of HF Issues
- Overall findings
- Training Issues

The projects interviewed will not be named for military and commercial confidentiality reasons.

Paper presented at the RTO HFM Symposium on "Advanced Technologies for Military Training", held in Genoa, Italy, 13 – 15 October 2003, and published in RTO-MP-HFM-101.

Approach Summary

In order to scope the project within the available budget, careful selection of projects had to be made to provide a representative sample of the platforms and equipment in service. This was carried out using a 4-dimensional criteria space covering project capability, size, time in service and level of procurement control. A semi-structured interview technique was determined to be appropriate for the assessment and a generic checklist was developed covering the full range of Human Factors topics. Each IPT received an outline of the topics that were going to be addressed during the interviews to ensure that the appropriate team members were available and fully briefed. The interviews were conducted with the aid of a recorder and the summary reports were sent to each IPT prior to the analysis being carried out. Further validation visits at service bases were carried out for about a third of the projects.

Project Selection

The down selection of the projects was based upon 4 criteria that were deemed to be significant following the first year of the study:

- Capability Area:
    - Strategic Deployment
    - Manoeuvre
    - Strike
    - Information Superiority

- Project Size:
    - Large – A, B (>£100M)
    - Small – C, D (<£100M)

- Length of Time in Service:
    - Mature – Post Mid life upgrade stage/long time in Service
    - Immature – Relatively short length of time in Service/Pre upgrade

- Level of Procurement Control
    - High – UK bespoke systems/high degree of specification control
    - Low Control – COTS, MOTS, upgrade of existing systems, international collaborative projects

It was recognised that the selection of 24 projects meant that there could not be a balanced sample of every criteria combination but it was agreed that there should be at least a split between high and low control and across all capability areas. The project selection within these categories is shown in table 1 below:

	Strategic Deployment (7)	**Manoeuvre** (6)	**Strike** (6)	**Information Superiority** (5)
Size 15 Large 9 Small	4 Large 3 Small	4 Large 2 Small	3 Large 3 Small	4 Large 1 Small
Maturity 10 Mature 14 Immature	5 Mature 2 Immature	2 Mature 4 Immature	2 Mature 4 Immature	1 Mature 4 Immature
Control 17 High 7 Low	5 High 2 Low	4 High 2 Low	5 High 1 Low	3 High 2 Low

Table 1 Summary of Numbers of Projects by Category

Taxonomy of Generic HF Issues

A checklist of HF issues was derived from a set of technical topics that had been successfully used on maritime projects (Sea System Publication SS10) and formed a good basis for discussion with the IPTs. The topics were:

- **Operational** – issues associated with operational procedures, role or scenario changes, task allocation etc
- **Organisation (Manpower)** – issues associated with manning (operation and support) and communication issues within and between teams etc
- **Personnel** – issues associated with user characteristics, skills, gender etc
- **Training** – issues associated with the effectiveness of training
- **Operability** – issues associated with the Human-Machine Interface affecting overall mission effectiveness
- **Layout** – issues associated with the physical layout of the working environment
- **Environment** – habitability issues (noise, heating, ventilation, lighting etc) that may affect overall human performance
- **Maintenance** – issues associated with equipment maintainability and availability
- **System Safety** – concerns about accidental damage to the user
- **Health Hazard** – concerns about human exposure to hazardous or unhealthy environments

In addition interviewees were invited to outline any additional topics/issues specific to the project/equipment that were not covered adequately by the topics/issues listed. The number and type of concerns raised reflected the knowledge and background of the interviewees and did not necessarily cover all of the topic areas. As most of the interviews were conducted within the Defence Logistics Organisation, where maintaining and sustaining capability is their primary focus, the concerns tended to be more support relate

Overall Findings

The issues were abstracted from the interviews and categorised within the 10 topic areas. Each topic area was further sub-categorised into root problem areas.

There were in total about 400 different concerns raised and the overall findings are shown here:

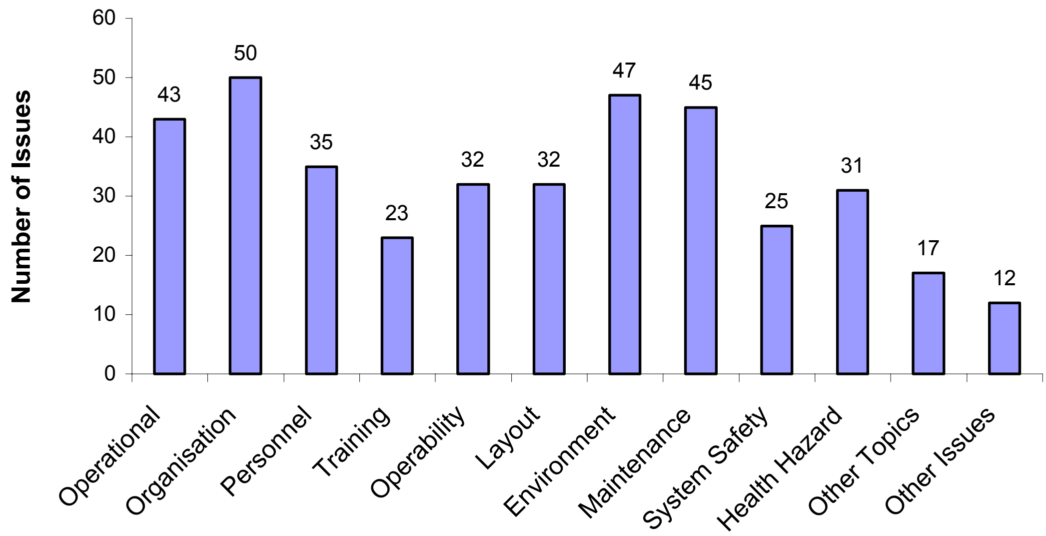


Figure 1 Number of Concerns within Topic areas

Each topic area was analysed against the highest number of reports of specific issues within each sub-category and table 2 below shows the top 3 issues within each of the topic areas:

Technical areas	Highest Issue and No of occurrences	2nd Highest Issue and No of Occurrences	3rd Highest Issue and No of occurrences	% of top 3 Issues Addressed
Operational	Role Change **11**	Environment Change **5**	Tactical Capability **5**	**62%**
Organisation/Manpower	Comms Capability **14**	Comms Quality **7**	Human Performance **7**	**68%**
Personnel	Personnel Availability **16**	Skill Level **8**	Physical Constraints **5**	**66%**
Training	Trainer Representation **6**	Maintainer Training Aids **5**	Trainer Utility/Utilisation **5**	**25%**
Operability	Control Operation **8**	Display Viewability **6**	Tactical Capability **5**	**58%**
Layout	Workspace **9**	Kit/clothing **5**	Operability Compromised **4**	**22%**
Environment	Temperature **14**	Noise **10**	Ventilation **5**	**28%**
Maintenance	Maintainance Access **11**	Availability of Spares **8**	Workload **7**	**27%**
System Safety	Unsafe Loads **6**	Physical Accident Risk **4**	Atmospheric Risks **4**	**86%**
Health Hazard	High Risk Conditions **8**	Toxic Substances **6**	Sensory Degradation **5**	**68%**
Additional Topics	Survivability **6**	Vulnerability **3**	Anthropometric Data Shortage **1**	**40%**
Additional Issues	Operational Effectiveness **4**	Can Do Attitude **3**	Configuration Control **1**	**38%**

Table 2 Summary of most frequent issues within each technical topic

The final column shows the percentage of concerns that were likely to be addressed during the life of the specific programmes. It can be seen from the table that the training concerns were one of the areas least likely to be addressed.

15 out of 24 of the projects highlighted specific concerns associated with training and the main areas of concerns were as follows:

- Trainer representation
- Maintainer training Aids
- Trainer Utility/utilisation
- Documentation support
- Training gaps and conflicts in training
- Scenario evolution

Together training facilities excluding documentation support amounted to over two thirds of the training issues raised. The diagram below shows the percentage of issues raised within each sub-category

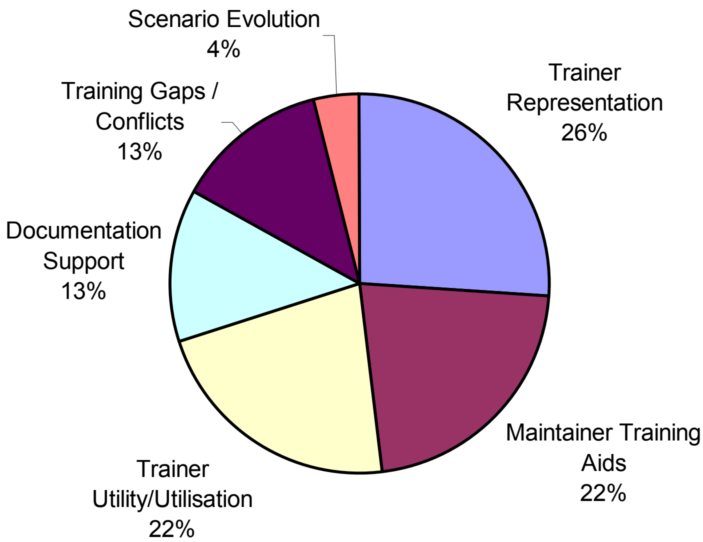


Figure 2 Relevant Training Issues

Trainer Representation

This area of concern was identified by 6 of the projects interviewed. Many of the training facilities appeared to be lacking in their operational representation and this increases the need for on-job training. The equipment was not flexible enough to give realistic training Thus in some cases operational readiness only comes from expensive training exercises at sea or where the fidelity is lacking, live firings on ranges. There were particular issues associated with older programmes where enhanced capabilities were being provided to the platforms but not to the trainers in the same timescales. The view from the projects interviewed was 'Front Line First' ie get the capability to the front line and let the training catch up. In some cases but not all, the lack of fidelity did have operational repercussions. The lack of visual or motion cues on the simulators also reduced the effectiveness of the training.

Maintainer Training Aids

5 of the projects interviewed cited poor training aids for the support staff. In some cases there were no trainers for the maintainers leaving all the skills to be acquired via on-job training. Generally the equipment care training was ineffective and one IPT provided photographic evidence of complete lack of equipment care – e.g. optics cleaned with a wire brush. Another outcome of poor maintainer training was the tendency

to change LRUs without proper fault diagnosis. This was largely attributed to a complex system where the technicians lacked adequate knowledge and skills to make the correct diagnosis.

Trainer Utility/utilisation

There were significant issues associated with trainer utility with 5 IPTs highlighting issues associated with the effectiveness of their training aids. In addition the use of the simulators to gain a much deeper understanding of the capabilities of the system potentially provided a valuable tactics development tool but poor utilisation meant that they were not making the best use of a highly complex and capable system. There were no airborne multiple crew simulators for the projects interviewed that enabled multi-crew platforms to train and rehearse effectively. However the naval platforms did have good training facilities, which were used very effectively. In addition the operational load and staff shortages made dedicated training time limited but generally on-job training was found to be effective.

Documentation support

3 of the projects interviewed highlighted problems associated with documentation support. The type of problems encountered varied between the obsolescence of the documentation media and total absence of any training material. In one particular case the media was a micro-fiche reader of which insufficient were available. Thus there was a high dependability on paper versions that were not always up to date. This had a particular impact upon maintenance training. In another case there was a complete lack of training material for the training support staff. Therefore they had to spend time with the manufacturers familiarising with the kit.

Training gaps and conflicts in training

The issues associated with training gaps and training conflicts were all associated with vehicle systems where drivers were both operators and maintainers. The training for drivers and maintainers was carried out at different establishments and this occasionally gave rise to conflicts and overlaps between the respective manuals that lay down crew maintenance and maintainer tasks. Doctrine and training occasionally had to be updated after incorrect use of procedures. Refresher training was highlighted as a problem by one IPT as each regiment has only one suite of gunnery simulators to serve 4 squadron of 14 crews. The amount of dismounted 'other duties' often led to long gaps without refresher training which could lead to 'skill fade'.

Scenario evolution

In one particular example the scenarios on the simulator were difficult to change and the technicians lacked the skill or training to make the necessary changes thus once the scenarios become predictable the training becomes ineffective. This directly impacted upon simulator utilisation as discussed above

Impact of Type of Project

As discussed previously the projects were selected against 4 criteria to determine whether the capability area, size of project, maturity of project and level of control influenced the numbers or types of concerns. The results are shown in the diagram below. The percentage bars show each category normalised as a percentage of the total number interviewed within that category in order to take account of distribution differences.

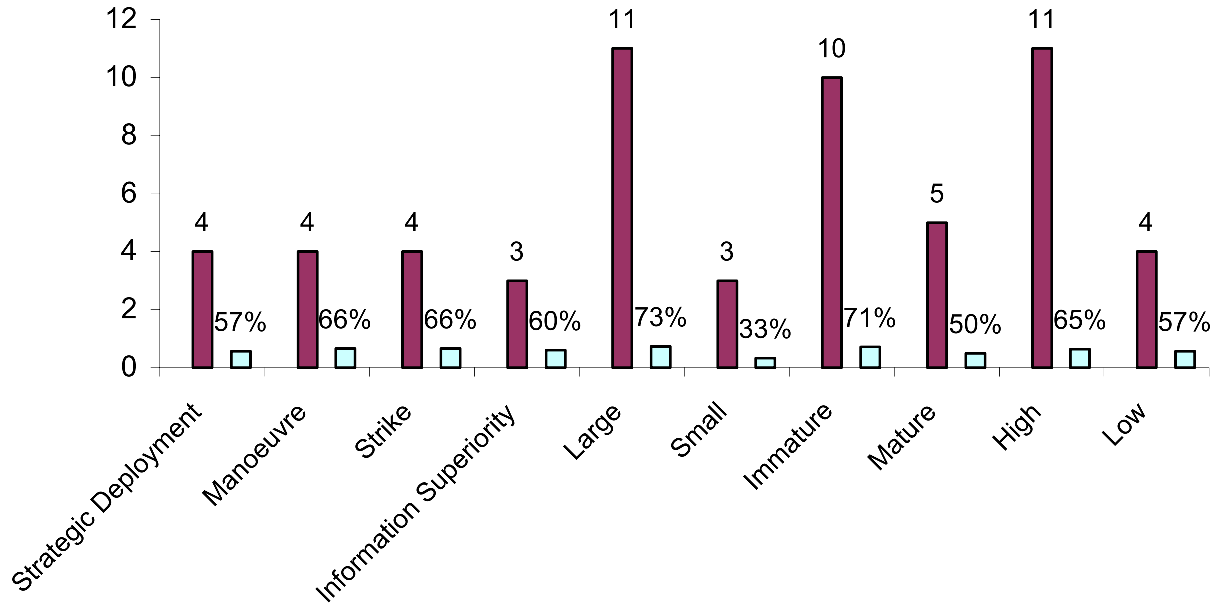


Figure 3 Summary of Issues against the project characteristics

This figure shows no significant differences in the number of training concerns across the 4 procurement areas but it does show that the majority of issues were those associated with large programmes within a pre-upgrade phase. The type of problem exhibited for immature programmes are mainly fidelity related. In addition the bespoke type systems (high control) yielded more problems than the COTS type systems (low control).

Conclusions

- Cost factors are driving the need for increased use of virtual training solutions. However the findings from the survey revealed that there is still a reliance on field-based training due to fidelity issues. These fidelity issues were affecting the utilisation of the simulators to the detriment of maximising the effectiveness of the systems.

- Achieving the right balance between simulator and field-based training continues to cause concern. Whilst the technology to create powerful graphic representations to give good visual feedback is available now and the processing capability enables this to be carried out in real time, it is likely that there will always be aspects of the task that cannot be trained in a simulator.

- Front Line First philosophy means that the upgrade to training devices may not provide the lead training time required. This could impact the operational effectiveness of the delivered capability.

- The advancement of highly complex IT systems is impacting upon selection and training of new skill areas. This may require novel approaches to training to acquire the level of knowledge and understanding of these complex systems.

- Lack of maintenance training aids is leading to ineffective and inefficient support to the front line. The focus on equipment care needs to be given more prominence.

- Maintenance of highly complex IT system trainers currently requires specialist technicians to ensure the systems can be fully exercised within the full range of threat scenarios. There are retention problems for these skill areas as there are lucrative contracts available within industry.

Recommendations

- The fidelity of the training device must not be compromised as this could result in negative training problems and impact upon operational effectiveness

- There needs to be greater consideration of embedded , on-job training facilities to compensate for shortage of personnel and off-job training time.

- Rapid augmentation of new capabilities into the training aids needs to be considered in the design of the simulators to ensure that they maintain the required lead training time.

- There needs to be much more emphasis on maintenance training and equipment care as this clearly impacts upon operational readiness of the equipment and platforms.

- The training of highly complex, multi modal systems needs further research to enable the level of mode awareness to be acquired although quite clearly the source of this problem is a design issue.

- Maintenance of simulators needs to be simplified to ensure ease of scenario update to enable full utilisation of the systems across a full range of scenarios.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO-MP-HFM-101 AC/323(HFM-101)TP/52	ISBN 92-837-1117-3	UNCLASSIFIED/ UNLIMITED

5. Originator
Research and Technology Organisation
North Atlantic Treaty Organisation
BP 25, F-92201 Neuilly-sur-Seine Cedex, France

6. Title
Advanced Technologies for Military Training

7. Presented at/Sponsored by
The RTO Human Factors and Medicine Panel (HFM) Symposium held in Genoa, Italy, 13-15 October 2003.

8. Author(s)/Editor(s)	9. Date
Multiple	April 2004

10. Author's/Editor's Address	11. Pages
Multiple	396

12. Distribution Statement
There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.

13. Keywords/Descriptors

	Distributed systems	Interoperability
Architecture	Distributed training	Learning management system
Augmented reality	Education	Professional military training
Communications networks	E-learning	Requirements
Compatibility	Embedded training	Research management
Computer based instruction	Emerging technologies	Standardization
Computer based training	Human factors engineering	Technology
Criteria	Information systems	Training effectiveness
Culture	Information transfer	Virtual reality
Distance learning	Integrated systems	Visualization technologies
Distributed learning	Internet technology	WWW (World Wide Web)

14. Abstract

Computer-Based Instruction, Training System Design, Distributed Training, and Virtual Reality are technologies that have dramatically changed. But many of the problems and issues related to training technologies, its application, acceptance, and effectiveness are still with us today. This unclassified Symposium theme was to share data and establish working relationships that can continue to deal with the problem of training and simulation from a human systems integration point of view.

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DIFFUSION DES PUBLICATIONS

RTO NON CLASSIFIEES

Les publications de l'AGARD et de la RTO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre est la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la RTO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (www.rta.nato.int) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Streitkräfteamt / Abteilung III
Fachinformationszentrum der
Bundeswehr (FIZBw)
Friedrich-Ebert-Allee 34, D-53113 Bonn

BELGIQUE
Etat-Major de la Défense
Département d'Etat-Major Stratégie
ACOS-STRAT – Coord. RTO
Quartier Reine Elisabeth
Rue d'Evère, B-1140 Bruxelles

CANADA
DSIGRD2
Bibliothécaire des ressources du savoir
R et D pour la défense Canada
Ministère de la Défense nationale
305, rue Rideau, 9e étage
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1, P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

ETATS-UNIS
NASA Center for AeroSpace
Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GRECE (Correspondant)
Defence Industry & Research
General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Centro di Documentazione
Tecnico-Scientifica della Difesa
Via XX Settembre 123
00187 Roma

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

PAYS-BAS
Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE
Armament Policy Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE
DIC Czech Republic-NATO RTO
VTÚL a PVO Praha
Mladoboleslavská ul.
Praha 9, 197 06
Česká republika

ROYAUME-UNI
Dstl Knowledge Services
Information Centre, Building 247
Dstl Porton Down
Salisbury
Wiltshire SP4 0JQ

TURQUIE
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320
ETATS-UNIS

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Acquisitions, Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants :

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
ETATS-UNIS

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
ETATS-UNIS
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

NORTH ATLANTIC TREATY ORGANISATION

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

RESEARCH AND TECHNOLOGY ORGANISATION

DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

AGARD & RTO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO reports, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your Organisation) in their distribution.

RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of RTO reports as they are published, please visit our website (www.rta.nato.int) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Etat-Major de la Défense
Département d'Etat-Major Stratégie
ACOS-STRAT – Coord. RTO
Quartier Reine Elisabeth
Rue d'Evère
B-1140 Bruxelles

CANADA
DRDKIM2
Knowledge Resources Librarian
Defence R&D Canada
Department of National Defence
305 Rideau Street
9th Floor
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC
DIC Czech Republic-NATO RTO
VTÚL a PVO Praha
Mladoboleslavská ul.
Praha 9, 197 06
Česká republika

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72
92322 Châtillon Cedex

GERMANY
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Friedrich-Ebert-Allee 34
D-53113 Bonn

GREECE (Point of Contact)
Defence Industry & Research General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HUNGARY
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ICELAND
Director of Aviation
c/o Flugrad, Reykjavik

ITALY
Centro di Documentazione Tecnico-Scientifica della Difesa
Via XX Settembre 123
00187 Roma

LUXEMBOURG
See Belgium

NETHERLANDS
Royal Netherlands Military Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

POLAND
Armament Policy Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide, P-2720 Amadora

SPAIN
SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

TURKEY
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

UNITED KINGDOM
Dstl Knowledge Services
Information Centre, Building 247
Dstl Porton Down
Salisbury, Wiltshire SP4 0JQ

UNITED STATES
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
UNITED STATES

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource locator:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
UNITED STATES

Government Reports Announcements & Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 2216
UNITED STATES
(also available online in the NTIS Bibliographic Database or on CD-ROM)

ISBN 92-837-1117-3