Study Report 2002-05

Individual and Collective Training in Live, Virtual and Constructive Environments

Training Concepts for Virtual Environments

Paul J. Sticha
Roy C. Campbell
C. Mazie Knerr
Human Resources Research Organization

United States Army Research Institute for the Behavioral and Social Sciences

April 2002

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

A Directorate of the U.S. Total Army Personnel Command

ZITA M. SIMUTIS
Acting Director

Research accomplished under contract for the Department of the Army

Human Resources Research Organization

Technical Review by
Billy J. Burnside, U.S. Army Research Institute
Bobby G. Witmer, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Study Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: TAPC-ARI-PO, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600.

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

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

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) April 2002	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) 2/9/00 – 3/31/01

4. TITLE AND SUBTITLE Individual and Collective Training in Live, Virtual and Constructive Environments – Training Concepts for Virtual Environments	5a. CONTRACT OR GRANT NUMBER DASW01-99-D-0012/006
	5b. PROGRAM ELEMENT NUMBER 665803
6. AUTHOR(S) Paul J. Sticha, Roy C. Campbell, & C. Mazie Knerr	5c. PROJECT NUMBER D730
	5d. TASK NUMBER 242
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Human Resources Research Organization 66 Canal Center Plaza, Ste 400 Alexandria, VA 22314	8. PERFORMING ORGANIZATION REPORT NUMBER FR-01-18
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Study Report 2002-05

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

Michael R. McCluskey, Contracting Officer's Representative

14. ABSTRACT *(Maximum 200 words)*:

The goals of this research are (a) to develop a method for evaluating the capabilities of virtual simulation to represent the tasks and missions within a military application domain, (b) to demonstrate the methods in two domains, and (c) to propose ways to integrate the method with existing doctrine. Initial activities surveyed existing training systems and reviewed the capabilities of key virtual environment technologies. From this survey, we identified capabilities most likely to impede successful development of virtual environment training systems. A review of existing methods of evaluating or predicting training effectiveness identified several candidates for incorporation into the method produced in this project. Based on the results of this review, we developed a method for Specifying Training Requirements in Virtual Environments (STRIVE), combining features from two existing methods. The STRIVE methodology assesses the capability of virtual environment technology to support task performance based on subject matter expert judgments of selected cues and responses needed to perform task activities. A demonstration of the model was developed using Microsoft Access97. The STRIVE methodology can be used during the concept exploration and definition phase of virtual environment training system design and can support the development of the Operational Requirements Document (ORD).

15. SUBJECT TERMS

Virtual Environment, Training Requirement, Task Performance Support (TPS) Code, Task Coverage, Simulation Capabilities

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	85	Michael R. McCluskey (407) 384-3990

Standard Form 298

FOREWORD

This work was completed as a part of the TRADOC Studies Program. The topic area for study was concerned with the application of virtual environments in training programs. Virtual environment technology has been applied successfully to train individuals, crews, and units to achieve and sustain proficiency in performing their missions. Previous work has investigated variables that influence performance in virtual environments as well as some of the factors that affect the training effectiveness of systems using virtual environment technology. Other studies have evaluated the ability of specific virtual simulations to represent and train military tasks. The results of these efforts have produced knowledge that could be used to guide decisions regarding the design, development, and use of current or future training systems.

However, this knowledge of the capabilities of virtual environment training systems has not been organized to produce specific methods to select virtual environment solutions for training requirements, to guide the design of such systems, or to estimate their training effectiveness. Methods used to evaluate existing training systems, such as the Task Performance Support (TPS) code (Burnside, 1990; SHERIKON, 1995), can enumerate strengths and weakness of systems, and provide guidance for future enhancements. However, there has been much less progress developing methods to guide the design and development of new systems employing virtual environment technology.

The goals of this study were (a) to develop a method for evaluating the capabilities of virtual simulation to represent the tasks and missions within a military application domain, (b) to demonstrate the methods in two domains, and (c) to propose ways to integrate the method with existing doctrine. The study findings were briefed to representatives of the TRADOC System Manager, Combined Arms Tactical Trainer (CATT) and Project Manager, CATT on February 28, 2001.

MICHAEL G. RUMSEY
Acting Technical Director

TRAINING CONCEPTS FOR VIRTUAL ENVIRONMENTS

EXECUTIVE SUMMARY

Requirement:

Virtual environment technology has been applied successfully to train individuals, crews, and units to achieve and sustain proficiency in performing their missions. However, knowledge of the capabilities of virtual environment training systems has not been organized to produce specific methods to select virtual environment solutions for training requirements, to guide the design of such systems, or to estimate their training effectiveness. The goals of this research are (a) to develop a method for evaluating the capabilities of virtual simulation to represent the tasks and missions within a given military application domain, (b) to demonstrate the methods in two domains, and (c) to propose ways to integrate the method with existing doctrine.

Procedure:

Initial activities surveyed existing virtual environment training systems and reviewed the capabilities of selected key virtual environment technologies. From this survey, we identified the specific capabilities that were most likely to be impediments to the successful development of a virtual environment training system. A review of the existing methods of evaluating or predicting training effectiveness identified several candidates for incorporation into the method produced in this project. Based on the results of this review, we developed a method for Specifying Training Requirements in Virtual Environments (STRIVE), combining features from two existing methods. A demonstration of the model was developed using Microsoft Access97. The demonstration focused on two sample problems, the Aviation Combined Arms Tactical Trainer – Aviation Reconfigurable Manned Simulator (AVCATT-A) and the Heavy Expanded Mobility Tactical Truck (HEMTT). The final activity addressed ways that the STRIVE methodology can be integrated into U.S. Army Training and Doctrine Command (TRADOC) policy on training design and development.

Findings:

The STRIVE methodology extends the existing methods to be applicable before a training system has been designed. The resulting procedure assesses the capability of virtual environment technology to support task performance based on subject matter expert judgments of selected cues and responses needed to perform task activities. The user of STRIVE selects the level of detail at which ratings will be made for each task, describes selected requirements of the training domain and design constraints, rates the cues and responses of the selected task elements, and assesses task step importance. Based on these input data, the method calculates a score representing the extent to which the task elements can be supported by virtual environment technology. The scores are migrated to lower levels and aggregated to higher levels up to the task level.

The feasibility of the procedure was demonstrated with two example problems from considerably different domains. Although the demonstration is not operational software, it represents all method functions and implements all selection procedures and calculations.

Use of Findings:

The STRIVE methodology can be used during the concept exploration and definition phase of virtual environment training system design and can support the development of the Operational Requirements Document (ORD). It can aid in the selection of the individual or collective tasks that are included in the operational requirements. The application of STRIVE can help ensure that the tasks assigned to virtual environment training are realistic given the current technological capabilities. Furthermore, STRIVE can help in the development of a coherent training strategy that coordinates training in live, virtual, and constructive environments

TRAINING CONCEPTS FOR VIRTUAL ENVIRONMENTS

CONTENTS

List of Tables

List of Figures

INTRODUCTION

Virtual environment technology has been applied successfully to train individuals, crews, and units to achieve and sustain proficiency in performing their missions. Previous research has investigated variables that influence performance in virtual environments as well as some of the factors that affect the training effectiveness of systems using virtual environment technology. Other research has evaluated the ability of specific virtual simulations to represent and train military tasks. The results of this research have produced knowledge that could be used to guide decisions regarding the design, development, and use of current or future training systems.

However, research knowledge of the capabilities of virtual environment training systems has not been organized to produce specific methods to select virtual environment solutions for training requirements, to guide the design of such systems, or to estimate their training effectiveness. Methods used to evaluate existing training systems, such as the Task Performance Support (TPS) code (Burnside, 1990; SHERIKON, 1995), can enumerate strengths and weaknesses of systems, and provide guidance for future enhancements. However, there has been much less progress developing methods to guide the design and development of new systems employing virtual environment technology.

Estimates of the training effectiveness of an existing system can capitalize on the experience of trainers or other subject-matter experts (SMEs) who are both knowledgeable about the training domain and familiar with the training system. Any uncertainties regarding system effectiveness can be clarified empirically, at least in principle. However, before a training system has been developed, effectiveness is much more difficult to estimate, due to the incomplete knowledge of SMEs and the impossibility of empirical evaluation. In this case, a behavioral analysis of the activities to be trained can aid the estimation process. In fact, Burnside (1990) expressed the utility for such an analysis for evaluating existing systems. One way to perform such an analysis is to replace holistic judgments of the adequacy of a system to meet a training need with judgments regarding categories of behavior that are linked to specific system components (e.g., the visual subsystem, terrain database, or motion component). However, this more detailed analysis multiplies the number of judgments required.

The requirement for a detailed description of training requirements may be met, in part, through formal job documentation. For individual activities, this documentation includes soldier manuals, equipment operator manuals, programs of instruction, and other descriptions of training programs. Documentation of unit activities is found in Army Training and Evaluation Program (ARTEP) mission training plans (MTPs) and drills, as well as field manuals. However, published documentation of tasks is often incomplete. For example, Ford and Campbell (1997) conducted a detailed analysis of a single armor brigade mission, “Conduct deliberate attack.” This mission covered six tasks and was described in approximately eight pages in the relevant ARTEP MTP. However, the detailed description of this mission produced by the analysis was about 20 times this length. Formal task documentation may also be out of date, particularly when doctrine has been recently changed or when the tasks involve new missions (e.g., Stability and Support Operations [SASO]), conditions, or equipment (e.g., changes resulting from digitization).

The rapid advancement of the capabilities of the technologies required to conduct virtual environment training affects the accuracy of any method to assess the training effectiveness that can be obtained using these technologies. Some advances come as a natural consequence of general improvements in processing speed, memory cost, and other evolutionary improvements in capability. It may be possible to predict the speed at which such changes occur with reasonable accuracy. However, in other areas, predicting the advancement in the capability of virtual environment technology requires assumptions about both the difficulty of the technological problems to be solved and the research funding allocated to solve them (see Jacobs, Crooks, Crooks, Colburn, Fraser, Gorman, Madden, Furness, & Tice, 1994). Because research may be funded by the military services, civilian government agencies, and private industry, the total amount of funding allocated to improve specific technological advancements is difficult to predict. Consequently, we have not addressed the problem of projecting future capabilities of virtual environment technology.

Project Goals and Activities

The research described in this report has the following three goals:

- Develop a method for evaluating the capabilities of virtual simulation to represent the tasks and missions within a given military application domain.
- Demonstrate the method in two domains.
- Propose ways to integrate the method with existing doctrine.

The basic question to be addressed in this project is the following: "Can a given task or set of tasks be adequately trained using virtual training technology?" The tasks to be trained may be performed by individuals, crews, or larger organizational units. Answering this question requires a comparison of the capabilities required to train the task with currently available virtual environment training technology and, consequently, requires the following three kinds of information:

- A characterization of the capabilities of relevant virtual environment training technology,
- A detailed description of the tasks that should be trained using this technology, and
- A methodology to compare the capabilities of technology with the requirements of the tasks, and to make an overall recommendation.

To be useful, the methodology must be integrated into the training development and management process, as specified by U.S. Army Training and Doctrine Command (TRADOC) Regulation 350-70 (1999).

The tasks conducted in this project provided the information necessary to design the methodology, develop examples of its application, and specify how it can be linked with existing training policy. Initial activities surveyed existing virtual environment capabilities used for armor and aviation training. This survey also included a review of the capabilities of selected key virtual environment technologies. From this survey, we identified the specific capabilities that were most likely to prove to be impediments to the successful development of a virtual environment training system.

A review of the existing methods of evaluating or predicting training effectiveness produced several candidates for incorporation into the method developed in this project. Based on the results of this review, we developed a method for Specifying Training Requirements in Virtual Environments (STRIVE). STRIVE combines features from two existing methods, TPS code analysis (Burnside, 1990; SHERIKON, 1995), and a model for Optimization of Simulation-Based Training Systems (OSBATS; Sticha, Blacksten, Buede, Singer, Gilligan, Mumaw, & Morrison, 1990). A demonstration of the model was developed using Microsoft® Access97.

In developing the demonstration, we focused on two sample problems. The two problems, based on the Aviation Combined Arms Tactical Trainer – Aviation Reconfigurable Manned Simulator (AVCATT-A) and the Heavy Expanded Mobility Tactical Truck (HEMTT), provide different challenges to the operation of the method. To complete these two examples, we identified sources of training task information, selected tasks to be included in the demonstration, and rated the selected tasks as required by the STRIVE method.

The final activity addressed ways that the STRIVE methodology can be integrated into TRADOC policy on training design and development. This activity focused on where in the process STRIVE could provide the greatest utility, and on the best sources for training task information.

Together, these activities work to develop and demonstrate an approach to evaluate the capabilities of virtual environment technology to represent the tasks and missions within a given military domain. The demonstration illustrates the operation of the method and the calculations employed in two domains, but does not represent a complete implementation of the method. Development of an operational version of the method was beyond the scope of this project, and would require additional work.

Outline of Report

The report begins by presenting background information to define the problem addressed by this research. This information describes virtual, constructive, and live simulation environments and presents the features that distinguish them. The report then reviews selected methods that could be used to evaluate or predict the training effectiveness of virtual environment training systems. It then continues with a description of the capabilities of virtual environment training, based on the survey of selected systems and other relevant literature.

Based on this review, two methods – the TPS code and parts of the OSBATS model – were chosen to be the core of the STRIVE methodology. STRIVE extends the TPS code by incorporating assessments of the selected activities that may be difficult to represent using virtual environment technology. Furthermore, it allows the user flexibility to determine the appropriate level of aggregation in which training activities should be evaluated – from the tasks to individual steps and performance measures. The report describes the method in detail and gives a guide to the operation of the demonstration system. The discussion of the demonstration highlights both its capabilities and its limitations.

Finally, the report discusses issues relevant to incorporating the method within the TRADOC training development and management process, and presents a summary and conclusions.

BACKGROUND

Emerging modeling and simulation (M&S) technologies have been applied to enhance training effectiveness and efficiency in several distinctly different ways. For example, in describing the advancements in tactical engagement simulation that had occurred in the previous 20 years, Gorman (1991) distinguished three type of simulation: (a) constructive simulation, consisting of aggregated computer models of military campaigns; (b) subsistent simulation, using actual military vehicles operating on instrumented maneuver ranges; and (c) virtual simulation, in which manned simulators interact in a synthetic battle environment. This distinction has been elaborated over the years and has been applied to training from individual to unit levels. Gorman's term, "subsistent" simulation has been replaced by the term, "live" simulation.

The distinction among different classes of M&S methods has been incorporated into Army training regulations, including TRADOC Regulation 350-70 (1999), which describes the *Systems Approach to Training Management, Processes and Products*. The following discussion briefly summarizes the distinctions among these three training environments, as specified by TRADOC Regulation 350-70.

Virtual Simulation

TRADOC Regulation 350-70 lists the following three characteristics in the definition of virtual M&S: (a) a replication of warfighting equipment and munitions, (b) a shared terrain database on which collective training can be conducted, and (c) potential links to live or constructive M&S. Virtual training, then, is "training executed using computer generated battlefields in simulators with approximate physical layout of tactical weapons systems and vehicles" (p. Glossary-37). An important characteristic of these trainers is that the terrain information is presented as a three-dimensional view that approximates the view seen from actual equipment, whether it is through the window or using other sensors.

Although the primary focus of this definition is on unit training for combat, consistent with Gorman's (1991) definition, the scope of virtual training is often considered to be much greater than that. In fact, Keesling, King, and Mullen (1998) consider only the replication of equipment in the definition of virtual simulation, and consequently enlarge the concept to include a variety of individual trainers that range from simple procedural trainers, such as panel trainers, to complex, full-mission, weapon system simulators. Collective simulations are included in this definition, as well, such as Simulation Networking (SIMNET) and the Close Combat Tactical Trainer (CCTT).

Future training applications of virtual environment technology will encompass additional missions that are further separated from the tactical engagements that were envisioned in the original definition. Such applications might involve maintenance training, training for humanitarian missions such as demining, medical training, or other requirements. The definition of virtual simulation and its distinction from other kinds of training in these new domains will need to be modified to incorporate the types of training simulations that are developed for these new applications. Alternatively, additional categories of simulation may be developed to describe training approaches used for these missions.

The wide variety of training devices that are incorporated into the definition of virtual simulation represents a range of technological capabilities. When there are no constraints on the types of virtual environment technology that can be used for a particular training device, the potential of the technology is best characterized by the capability of the most advanced individual, crew/team, and collective training systems. However, design constraints regarding size, cost, or compatibility with other systems can limit the potential capability of virtual environment technology. Because this project is concerned with the potential of virtual environment technology, it focuses on advanced training systems. However, it considers some of the constraints that may limit the potential of the technology.

A final aspect of the definition of virtual M&S is that it may include links to live or constructive simulations. This capability can blur the distinction between these three kinds of models and simulations in ways that will be discussed under the heading of hybrid simulations.

Constructive Simulation

TRADOC Regulation 350-70 defines constructive M&S as "Models, simulators, and/or simulations that involve real people making inputs into an M&S entity that carries out those inputs by simulated systems" (p. Glossary-11). Constructive simulation is generally used to exercise unit command and staff functions at any echelon.

In a similar fashion to virtual simulation, constructive simulation can vary over a wide range in sophistication. For example, Keesling et al. (1998) include in their definition of constructive simulation non-automated simulations, such as sand tables and terrain maps, in addition to the more commonly considered computer wargaming models, such as the Battalion Brigade Simulation (BBS) and Janus.

In constructive simulations, the actions of individual vehicles and weapon systems are simulated, as are the results of engagements. For unit commanders who are normally located in a vehicle, such as armor platoon leaders, constructive and virtual simulation are quite different and stress different aspects of command tasks. However, for commanders and staff who are normally located in a tactical operations center (TOC) using maps, computers, and communication devices, there is essentially no difference between stimuli presented by constructive and virtual simulation methods (and indeed, live simulation).

Live Simulation

Live training consists of "training executed in field conditions using tactical equipment, enhanced by training aids, devices, simulators, and simulations (TADSS) and Tactical Engagement Simulation (TES) to simulate combat conditions" (TRADOC Regulation 350-70, p. Glossary-20). As Keesling et al., (1998) point out, there are two types of live simulations: (a) live fire exercises (LFXs), in which participants fire full-service ammunition at targets on ranges; and (b) force-on-force exercises, in which live forces interact using instrumented weapon systems, such as the Multiple Integrated Laser Engagement System (MILES).

An LFX can provide realistic training for both soldier and collective skills. This training can reinforce tactical skills and develop a soldier's confidence in himself, as well as in teammates, leaders, and equipment (Burkett & Mullen, 2000). In fact, Burkett and Mullen have argued that in live fire training, "more is learned than that which is identified by the standards, task steps, and performance measures reflected in the mission training plan (MTP)" (p.11). However, conducting an LFX requires a considerable investment in maneuver space, equipment, and ammunition. Consequently, these exercises are usually limited to the platoon or company team level.

Force-on-force exercises, also referred to as live M&S, are typified by the combat training centers (CTCs). These exercises provide the most realistic peacetime environment possible for combat training. The level of realism for the interaction between troops in live M&S is not possible with live fire exercises. Army Field Manual (FM) 25-101 (1990) portrays both the high level of realism and the high resourcing required for live M&S, both of which exceed LFXs.

Hybrid Simulations

The fact that virtual simulations may include links to live or constructive simulations opens up the possibility for the existence of hybrid simulations that have aspects of more than one simulation category. In fact, an important constructive component of collective virtual simulators, such as SIMNET and CCTT, is the semi-automated forces (SAF) that represent opposing forces, adjacent units, or echelons that are not represented by live soldiers. The SAF respond to the general guidance of a controller, but their specific actions are determined to a great extent by conditions of the environment, the terrain, and the proximity to friendly and threat vehicles. This type of behavior fits the definition of constructive simulation described previously. The existence of constructive components is common in individual, crew/team and collective virtual simulations.

A much greater degree of integration between virtual, constructive, and live simulations is represented by the concept of the synthetic theater of war (STOW). This concept envisions a training and mission-rehearsal environment in which units participating in all three types of simulation would work together to perform a mission. Participants would interact as if all were on a common battlefield, even though some would be operating on actual terrain, others in a virtual environment, and still others would be computer-generated.

REVIEW OF SELECTED ANALYSIS METHODS

Considerable research and analyses have been conducted over the past 20 years to establish the training effectiveness and cost effectiveness of various training system designs. Because there are recent reviews of this research (see Muckler & Finley, 1994; Simpson, 1995), our summary will focus on the methods that are specifically related to the project goals. That is, the methods covered in this report could be used to determine the extent to which a set of training requirements could be met using a given technology.

Our summary and review of existing methods focuses on the following questions that relate to the appropriateness of these methods for this project.

- What is the source of training requirements and what is the level of detail at which they are represented?
- What kind of information about the training requirements is used to make the evaluation? How much and what kind of information is provided by subject-matter experts (SMEs)?
- Does the method apply to existing training systems only, or to future systems?
- What characteristics of training systems are considered by the method? Are they considered as a unit, or are they subdivided into subsystems or components?
- What is the basis of the elemental assessment of effectiveness? How are assessments at the elemental level aggregated to obtain an overall measure of effectiveness?
- What kind of a recommendation is made by the method? How can this recommendation be used in the training system development process?

All of the methods reviewed in this section require input from SMEs who are able to understand the training requirements and to interpret the implications that these requirements have on the training system capability needed to meet them. The role of the SME is critical, given the incomplete description of activities found in formal documentation. However, one must recognize that the knowledge of the SME is also necessarily limited by experience with the training domain and familiarity with the capabilities of training technology. Also, methods differ regarding the number of judgments that they demand from the SME. The sheer number of judgments may tax the SME's attention and ability to discriminate.

The goals for this project dictate the need for a method that applies to future systems and requires a reasonable number of judgments by SMEs in the areas of their expertise. Thus the answers to the preceding questions will determine which methods can be used, either alone or in combination, to evaluate the ability of virtual environment technology to meet a set of training requirements.

Task Performance Support (TPS) Codes

The TPS code provides a straightforward method to evaluate the ability of a training system to support the performance of tasks that must be trained. This method was developed by

Burnside (1990), who used it to evaluate the capabilities of the existing version of SIMNET and to identify those enhancements that would improve the training effectiveness of the system. SHERIKON (1995) modified the method and applied it to evaluate the capabilities of CCTT. Most recently, the TPS code was used in the CCTT Accreditation Report (1999). These versions vary somewhat in detail but share the same general approach. We present Burnside's original method in some detail, and then describe how the later methods have modified the original approach. The differences between these methods are summarized in Table 1.

Table 1
Summary of TPS Code Methods

Method	Rating Scales for Performance Measures	Aggregation Methods	
		Task Steps	Tasks
Burnside (1990)	Five levels of training support: Highly Partially Minimally Outside support required Not supported	Combine to same five levels of support based on: 1. Number of standards in task step 2. Percentage of standards with each level of rating	Combine to same five levels based on: 1. Number of task steps in task 2. Percentage of task steps with each level of aggregated rating
SHERIKON (1995)	Two kinds of ratings 1. Four levels of training support: Highly Moderately Outside support required Not supported 2. Not Critical – can be supported but is not critical to the task step	Combine to same four levels based on 1. Number of performance measures in the task step and 2. Percentage of performance measures with each level of rating 3. Criticality rating does not enter the combination	Combine to five numeric levels based on 1. Number of task steps in the task 2. Percentage of task steps with each level of aggregated rating 3. Exception when one task step; only three levels are used
CCTT Accreditation Report (1999)	Two kinds of ratings 1. Four levels of training execution (support): High Moderate Low No Support 2. Feedback ratings have the same four levels	1. Combine to same four levels of support and feedback (separately). 2. Convert to numeric scale 3. Assess four levels of importance to train in CCTT: Essential Medium Low Not Important	1. Weight task step aggregates by multiplying them by importance ratings. 2. Calculate TTCF on a normalized scale (separately for support and feedback). 3. Arrange task rankings in quartiles (High, Moderate, Marginal, and Low)

The Basic Method

The basic method that Burnside (1990) developed started with ratings of the ability to support ARTEP MTP performance measures using SIMNET. He rated the support at the level of the performance measure[1], which provides greater detail than task steps, to render more accurate estimates of effectiveness. He then used a rule-based method to combine the performance measure ratings to obtain summary ratings at the task step and task level. The rules for combining the detailed ratings of performance measures were algorithmic and thus the consolidation could be automated.

The descriptions of training requirements that were used to evaluate the effectiveness of SIMNET were taken from the following three Armor ARTEP MTPs:

- ARTEP 17-237-10 MTP for the tank platoon,
- ARTEP 71-1 MTP for the tank and mechanized infantry company team, and
- ARTEP 71-2 MTP for the tank and mechanized infantry battalion task force.

The task descriptions in these documents state the general conditions for performance and the overall standard for successful performance. The task steps within a task are listed sequentially. Each task step has detailed performance measures, which specify measurable activities that must be conducted or outcomes that must be achieved for the task step to be performed correctly.

Burnside developed the following scale for rating the extent that SIMNET could support performance specified by the detailed ARTEP MTP performance measures:

- Highly Supported (H) – The performance measure can be supported entirely. All required actions can be performed in essentially the same way that they are in field training or combat.

- Partially Supported (P) – The performance measure can be supported to a large extent. The majority of required actions can be performed realistically, while the remainder can be performed somewhat artificially due to the limitations of the system.

- Minimally Supported (M) – The performance measure can be supported to a limited extent. The majority of the required actions must be performed under artificial conditions, although some may be performed realistically.

- Outside Support Required (O) – The performance measure can be supported in the simulation facility, but the majority of actions must be performed outside of the simulation.

---

[1] Burnside used the terms "standard" and "subtask" to refer to the elements of tasks that are termed "performance measure" and "task step" in more recent ARTEP MTPs. The more recent terms were also used in applications of TPS Code analysis by SHERIKON (1995) and in the CCTT Accreditation Report (1999). To be consistent with current naming conventions, we use the latter terms, both in the discussion of the TPS Code and in the description of the method developed in this project.

- Not Supported (N) – The performance measure cannot be supported using the system or in the facility. A significant portion (more than 25%) of the required actions cannot be performed either using the simulation or within the facility using outside support.

Burnside, who had knowledge of both the actions required to perform each task and the capabilities of the simulator to support these actions, made these ratings. These ratings for the 4,381 performance measures addressed in the evaluation were then aggregated to obtain summary ratings at the step and task level. The aggregation rules used to derive task step ratings from ratings of performance measures considered two factors: (a) the number of performance measures included in each task step and (b) the distribution of the ratings of the performance measures within each task step. The rules were nonlinear and produced task step ratings on the same five-point scale that was used to rate the performance measures.

Task ratings were obtained from task step ratings using similar, nonlinear aggregation rules. In addition to the number of steps and their ratings, the task aggregation rules considered the criticality of the task steps to determine an overall task rating. Criticality of performance measures could not be considered, because that information was not provided in the ARTEP MTPs. The task assessments were expressed on the same five-point scale that was used to assess performance measures and task steps.

Only 35% of the ARTEP MTP tasks were rated as at least partially performable in SIMNET. Of the three echelons rated, the platoon had the highest percentage of trainable tasks, but also had the highest percentage of tasks that were not trainable. Burnside noted that this result was not a criticism of SIMNET, because it was not designed to train all of the collective tasks.

Burnside suggested that the method be used for training development or testing of training strategies. For example, unit leaders could use it to decide which tasks to train in the simulation and which in field exercises. The detailed information could support training strategy plans; for example, if the unit leader wanted to train a task in the simulation that is only partially supported, the training plan could give emphasis during field training to the task steps that the simulation does not support. An additional use is in operational testing, to select tasks for training effectiveness and transfer studies.

Variations on the Method

Both SHERIKON (1995) and the CCTT Accreditation Study (1999) used assessment methods based on Burnside's (1990) proceure. Each of these analyses made some changes to the original method. We briefly present the modifications that were made in each case.

In their analysis of CCTT capabilities, SHERIKON (1995) modified both the rating scale used to assess performance measures and the rules used to generate task step and task summary ratings (see Table 1). The developers wanted to minimize the number of levels in order to make the assessment process easier for the SMEs making the judgments. Consequently, they substituted a single category, labeled "moderately supported," for the categories, "partially supported" and "minimally supported," used by Burnside (1990). The developers added another category that is unrelated to support, but assesses the criticality of the performance measure to the task step. This category labeled "not critical" indicates that the performance measure can be supported by the simulation but is not critical to the training of the associated task step. This category appears at the performance measure level only and is not used to rate task steps or tasks.

Two major changes were made to the aggregation rules. First, the rules for combining performance measure ratings to task step ratings were modified slightly to reflect the changes in the performance measure rating scales. The rules appear to be slightly more stringent than those used by Burnside (1990), requiring a greater percentage of the performance measures to receive the highest rating for the entire task step to receive that rating. Second, the task performance scale was made a 5-point numeric scale taking values from 0 to 4. The highest level indicated that all task steps were rated as "highly supported." The lowest level indicated that fewer than 30% of task steps received a rating of "highly" or "moderately supported."

A final change that was made by SHERIKON (1995) was that performance measures were rated for both day and night conditions. Separate ratings were maintained and were not combined at any level of the assessment.

The CCTT Accreditation Report (21 May 1999) included a TPS code analysis as one element of a more extensive evaluation. This analysis also followed the general guidelines developed by Burnside (1990), but the details of the method were different from both of the previous analyses. Ratings were applied to performance measures using a scale consisting of the following four categories: High support (H), moderate support (M), low support (L), and no support (N). Performance measures were rated separately concerning the ability of CCTT to support the execution of the task and the ability of CCTT to provide appropriate performance feedback. Execution ratings assessed whether the simulation provided the appropriate cues and/or whether it detracted from the cognitive processes.

Task step ratings were then calculated from the performance measure ratings using the same heuristic procedure used by SHERIKON (1995), adjusted to reflect the change in the names of the response categories. Once these task step ratings were determined, they were converted to numeric scores on a four-point scale from 0 (no support) to 3 (high support).

Task ratings were then calculated as a weighted sum of the task step ratings. Criticality weights were assessed for each task step by TRADOC representatives. These weights were then normalized to range between 0.0 and 1.0 and multiplied by the task step ratings to obtain a weighted rating for each task step. The weighted ratings were then summed over task steps to obtain a task rating, which was normalized to be on a 100-point scale. This overall value was termed the Training Environment Task Contribution Factor (TTCF). In addition, for critical task steps, a task step rating of zero mandated that the associated TTCF be zero.

The evolution of the TPS code has shown a movement from qualitative evaluation to numerical assessment scales. In addition, in the CCTT Accreditation Report, a nonlinear heuristic rule was replaced by a linear algebraic rule. Linear combination rules often provide reasonable approximations of more complex rules in situations in which assessments are imprecise or apt to contain error. Thus, the use of these rules seems appropriate in this case. On the other hand, the numerical scales are primarily used as a convenience; they do not represent equal intervals of training effectiveness.

Need for Enhancement of TPS Codes

Although the versions of the TPS code differ in detail, they share properties that may lead to difficulties in applying them to assess the effectiveness of virtual environment technology. These problems come from the assessment load that they place on the SME, their applicability to training systems that have not been developed, and the need to incorporate behavioral analyses.

SME Rating Workload

The TPS code places a high workload on the SME because of the sheer number of assessments required and because the assessments require knowledge of both the tasks that must be performed and the capabilities of the training system. Table 2 shows the number of tasks, task steps, and performance measures in the three applications of TPS codes that were reviewed. The number of performance measures rated in the CCTT Accreditation report was not reported, but we may estimate from the number of tasks and task steps that this number is somewhere between the numbers for the other analyses, perhaps about 6,400. Because both the SHERIKON analysis and the CCTT Accreditation Report required two ratings per task step, the total number of judgments is twice the number of performance measures – between 12,800 and 16,300 for these analyses.

Table 2
Number of Tasks, Task Steps, and Performance Measures Rated in TPS Code Analyses

Source	Tasks	Task Steps	Performance Measures
Burnside (1990)	182	1,095	4,381
SHERIKON (1995)	329	1,955	8,157
CCTT Accreditation (1999)	219	1,575	6,432 (estimated)

Whenever an individual or group of individuals makes so many judgments, it is reasonable to ask whether those judgments represent independent pieces of information, or whether they could be summarized by a relatively small number of simpler heuristics. A thorough evaluation of this question would require detailed interviews with the SMEs, but it is possible to get some understanding by looking at the variability of judgments across performance measures within a task step, and across task steps within a task. Differences between performance measures directly reflect differences in the judgments of the raters. Differences between task steps reflect both differences in the ratings and the effects of the rules that are used to combine ratings of performance measures to produce task step summary ratings.

Table 3 shows the maximum difference between the ratings of performance measures within a task step, and between task steps within a task. A difference of zero indicates that all performance measures (or task steps) within a task step (or task) received the same rating, while the maximum rating of three or four (depending on the rating scale used) indicates that the entire rating scale was used. The table indicates that for roughly one-half of the task steps, there was no difference in the ratings of the performance measures contained in them. Since there are at least four times as many performance measures as there are task steps, a procedure that used judgments at the task step level for the task steps with no variation in performance measures could reduce the number of ratings required by as much as 40%.[2] However, direct ratings at the task step level may be more difficult than at the performance measure level, and time and effort would be required to determine whether ratings should be made at the performance measure level or the task step level. Consequently, it is not clear how much time would be saved by making some ratings at the task step level.

Table 3
Distribution of Maximum Rating Difference within Task steps and Tasks (percent)

Source	Maximum Rating Difference				
	0	1	2	3	4
Performance measures within Task steps					
Burnside (1990)	44	19	14	7	16
SHERIKON (1995) – day, night, and total ratings[3]	65	8	4	23	
	55	17	8	20	
	54	18	4	24	
Task steps within Tasks					
Burnside (1990)	19	22	17	14	28
SHERIKON (1995) – day, night, and total ratings	40	25	15	20	
	43	23	17	17	
	35	30	15	20	
CCTT Accreditation (1999) – capability, feedback, and total ratings	38	33	15	14	
	32	15	40	12	
	27	12	44	17	

The task-level data shown in Table 3 indicate greater variability. This variability is particularly noticeable in the Burnside (1990) ratings, in which task steps with a task received the same rating for only 19% of tasks. Consequently, it appears that because of the variability in the activities included within a task, direct ratings cannot be made at this level.

New Training Systems

A second aspect of the workload that TPS code analysis places on the raters is the detailed knowledge that is required regarding both the activities that must be accomplished to perform the unit tasks and the capabilities of the simulation to provide a suitable environment in which to perform these tasks. The procedure requires the rater to be knowledgeable about both

---

[2] This estimate is an upper bound because it does not consider differences in ratings of performance measures between task steps containing more or fewer performance measures. There would be no savings for task steps that contain a single performance measure (14% of task steps in Burnside's data). An analysis that looked at maximum rating difference as a function of number of performance measures was not performed.

[3] The data for the SHERIKON (1995) assessment address ARTEP 17-57-10-MTP for the Scout Platoon only. Rating data for other units were not available for this analysis.

the tasks and the training system being evaluated. This requirement limits the usefulness of the method in the design phase before the training system has been developed, because the raters cannot have any experience with the system at the time.

The previous remark should not be interpreted as implying that TPS code analysis has no value before a training system has been developed. New training systems often provide evolutionary improvements over existing systems. It may be possible for raters to rate a new system in the design phase by considering a similar existing system and adjusting the ratings to reflect the incremental improvements included in the new system. For example, it is possible to use an analysis of SIMNET to make judgments about the effectiveness of CCTT to train the same tasks (Burnside, 1990). However, when the improvements are extensive or there is no precursor system, TPS code analysis might not be feasible.

Behavioral Analysis

Burnside (1990) suggested incorporating behavioral analysis into the method to improve its accuracy. Behavioral analysis can provide other benefits as well. The cues and response feedback requirements derived from a behavioral analysis could provide information specifying the simulation capabilities required to train activities in a virtual environment. These requirements, in turn, could be used to evaluate systems that are being designed, as well as existing systems. The benefits of behavioral analysis come at a cost in both effort and time to the SMEs who must provide the judgments that form the basis of the analysis. To be feasible, the behavioral analysis must be efficient and must focus on the information that will be most diagnostic in evaluating the sufficiency of a virtual simulation to train the activities.

Methods to Supplement TPS Codes

Several approaches have incorporated elements of behavioral analysis into the design and evaluation of training systems. Although a complete review of these methods is beyond the scope of this report, we highlight some recent and promising approaches. In addition, a method to conduct comparison-based predictions will be summarized. This method was specifically designed to evaluate a training system design before the system was developed.

Behavioral Methods

Behavioral analysis methods vary in the amount of information they obtain and regarding whether the same information is obtained for each activity being analyzed. Perhaps the simplest of these methods was used by Keesling, King, and Mullen (1998), who considered two types of learning associated with training requirements: (a) cognitive learning and (b) psychomotor learning. This information was then used to match training requirements with constructive, virtual, live, and hybrid training environments. The overall model for performing the match was based on the Automated Instructional Media Selection Model (AIMS; Kribs, Simpson, & Mark, 1983). In addition to the type of learning, AIMS considered the nature of performance cues, the types of responses, the ways that performance should be evaluated, the level of learning required, and special needs (e.g., memorization or crew interaction). Keesling et al. modified the AIMS model to consider cues, responses, evaluation, and three levels of learning.

The results of the analysis indicated that cognitive tasks could be trained well in all four environments, although the match was highest for a live environment and lowest for a constructive environment. Psychomotor tasks could not be trained well in a constructive environment, but could be trained equally well in virtual and live environments. In this way, the simple characterization of training requirements produced a modest distinction between training environments. Keesling et al. combined the information from this analysis with notional cost information to produce guidelines regarding the preferred training environment as a function of echelon, type of learning, and skill level.

Keesling et al. extended the results of the previous analyses to produce an Environment Selection Decision Aid (ESELDA) that provides a framework for assessing characteristics of training environments as they may be affected by the introduction of new combat systems. It is intended to be used by high-level decision-makers during early stages of the procurement cycle when the training environment and possibly the combat system have not been fully specified. The model evaluates the four training environments – live, virtual, constructive, and hybrid – on the following four characteristics:

- Feasibility. This factor assesses the effects of any impediments to implementation such as resources, technical challenge, or unacceptable risks.
- Affordability. This factor encompasses all lifecycle costs, as well as cost-related factors such as relative efficiency, impact on legacy systems, accommodation of future systems, and creation of hybrids.
- Suitability. This factor assesses the degree to which functions and tasks can be trained in the environment, considering cues, support for appropriate responses, incorporation of psychological stress, provision for feedback, and support for part-task training. The authors consider TPS codes to be an appropriate methodological framework for assessing suitability.
- Deployability. This factor assesses the ability of the environment to be deployed to units, considering both the logistical burden of deployment and the requirement for support of the environment in a deployed situation.

The overall value of a training environment was assessed as the weighted average of ratings on each of the four factors.

An intermediate level of behavioral detail is used by the Device Effectiveness Forecasting Technique (DEFT; Rose, Martin & Yates, 1985; Rose, Wheaton & Yates, 1985a, 1985b). DEFT is designed to assess the effectiveness of training devices at any stage in the development process. DEFT includes procedures to address the following four sets of questions:

- What is the performance deficit? That is, how much do the trainees need to learn to meet the performance criterion? How difficult are the required skills and knowledges to learn?
- What kind of features does the training system possess that will make learning more efficient?

- What is the residual performance deficit after use of the training system? How difficult is it to learn the skills necessary to meet operational performance objectives? What are the physical and functional similarities between the training system and the operational equipment?

- What is the anticipated transfer efficiency, given the training principals and instructional features used by the system?

Three different versions of DEFT allow these questions to be assessed at different levels of detail, from the overall program level to the task step level.

Embedded in the DEFT models are principles of learning that form the heart of the evaluation. These principles address such issues as techniques for learning long procedures, effects of overlearning on retention, use of memory aids, knowledge of results, and use of augmented cues early in training. The information contained in these principles specifies a form of behavioral analysis that can improve the assessment of effectiveness of virtual environment training technologies.

The method of behavioral analysis that provides the greatest detail and is the most relevant for the determining the ability of virtual environments to support training requirements was developed as a part of the larger decision support for the Optimization of Simulation-Based Training Systems (OSBATS; Sticha, Blacksten, Buede, Singer, Gilligan, Mumaw, & Morrison, 1990). Two sets of rules were developed to determine fidelity and instructional feature requirements on a task-by-task basis. The results of these rules supported the analytical procedures used by the other elements of the OSBATS model. Although these rules primarily considered the requirements for rotary-wing flight simulators, the procedures have more general applicability.

The fidelity requirement rule base uses task data to assess the types of cues that must be presented in order to train the task in a simulated environment. It uses backward chaining so that it asks for the minimum amount of information needed to make a recommendation. It considers 11 dimensions that describe visual, aural, and acceleration cues (See Table 4). To illustrate the operation of the rule base, we will describe how it makes recommendations regarding the required resolution for the visual display of the battlespace. Some of these rules are specific to rotary wing operations, but others can generalize to other mission areas.

The fidelity rule base considers five activities that may require high visual resolution to be performed in a simulated environment: (a) detecting small or distant objects, (b) estimating altitude, (c) estimating slant range to an object, (d) judging clearance between the aircraft and a nearby object, and (e) landing on a slope. If a task requires any of these activities, then the resolution required to perform that activity is determined. For example, if the task requires the soldier to detect small or distant objects, then the minimum size of the object to be detected and its maximum distance are obtained from the SME. The resolution requirement can then be estimated from the angle subtended by the object at its maximum distance. In a similar manner, the resolution required to estimate altitude can be obtained from the altitude that must be estimated, the tolerance with which the altitude must be maintained, and the size and distance of the objects that are used to make the estimation. The rule base only obtains data for activities that are relevant for a particular task, so that the data requirement is minimized. Furthermore, if a visual system representing the terrain is not required, the requirement for all visual variables is set to zero.

Table 4
Cue and Response Attributes from the OSBATS Model

Simulation Dimension	Range of Values	
	Low	**High**
	Image Generation	
Database Size	5 km x 5 km	30 km x 40 km
Visual Content	Plane with scattered trees	High-density hydrographic features (urban environment)
Visual Texture	Texture using modeling elements (lines and polygons)	Many digitized photos for texturing
Special Effects – Points	None	Cultural lights, weapons blast, damaged vehicles, airborne vehicles, moving ground vehicles
Special Effects – Area	None	smoke, dust, rotor wash
	Visual Display	
Visual Resolution	Sufficient to detect m^2 at 300 m (approx. 12 arc min/optical line pair[OLP])	Sufficient to detect m^2 at 4000 m (approx. 1 arc min/OLP)
Front Field of View (FOV)	40° x 40°	40° x 60°
Side FOV(s)	1 side, 40° x 40°	2 sides, 50° x 60°
	Motion Cueing	
Platform Motion	None	6 Degrees of Freedom (DoF)
Seat Motion	None	Seat Shaker and G-Seat
	Other Components	
Audio Effects	None	Weapons, Skid, Failures, Normal and Abnormal Operations Noises
Instructional Support Features	None	21 instructional support features

Comparison-based Methods

A new training system often has many characteristics of existing systems. Consequently, it might be possible to estimate the training effectiveness of a new system by comparing it to one or more similar existing systems. Comparison-based prediction methods (Klein, 1982; Klein, Johns, Perez, & Mirabella, 1985) provide a formal process to design proper comparison cases, identify the causal factors or high drivers that determine effectiveness, obtain appropriate information from SMEs, and document the procedure and resulting effectiveness estimates.

Using this approach, the SMEs are required to make comparative judgments rather than absolute judgments, a process that should improve the accuracy and reliability of the judgments.

In a test of the comparison-based prediction method, Klein (1982) applied the procedure to evaluate two training devices developed under the Army Maintenance Training and Evaluation Simulation System (AMTESS) project. SMEs compared them to the devices currently in use to train troubleshooting for automotive engine starting and operation. SMEs were told to assume two hours of training on the device and to estimate the time saved by using the device compared to training on the actual equipment.

Results showed that the scatter of estimates was wider for the device that had only paper documentation than for the device that was in use on-site with the SMEs. Klein concluded that the technique could not be reliable if the description of the device was poor; obtaining an adequate device description will be a problem in early design phases of acquisition. The need for a description of the training program that will use the device was also cited as a problem. Finally, selection of appropriate SMEs was a problem, because few people had enough experience to be judges. Although the evaluation of comparison-based predictions uncovered some problems with the method, elements of the method may be useful additions to the overall methodology.

Summary of Methods

The TPS code analysis describes a process that produces the type of information that is desired in this project. That is, it determines the extent to which a given technology can meet training requirements. Certain aspects of this approach, such as using ARTEP MTPs to specify unit training requirements and employing aggregation rules to derive summary assessments from more detailed ratings, should be at the core of any method to evaluate the efficacy of virtual environment technology. However, the requirements this method places on the rater and the knowledge it requires make it especially difficult to apply for training systems that are currently being designed.

The behavioral analyses and comparison-based methods that were identified do not by themselves address the problem at hand. However, some of these methods could be combined with a form of TPS code analysis to produce the required solution in a way that avoids some of the problems of TPS codes. The OSBATS rule base seems to be the most reasonable method to supplement the capabilities of TPS codes for several reasons. First, the task orientation of the method is consistent with the orientation of TPS codes. Second, use of the rule base reduces the knowledge required of the SME by eliminating the need for detailed knowledge of the capabilities of the training system. Third, the rule base is organized to minimize the number of judgments required to assess simulation requirements.

CAPABILITIES OF VIRTUAL ENVIRONMENT TECHNOLOGY

The process of evaluating an existing training device is facilitated by the fact that the technological capability of the device is relatively fixed and can be well understood by the individuals performing the evaluation. Evaluating the capabilities of a set of technologies in the absence of a specific training device requires an understanding of relevant components of the technology and their capabilities. Specifically, the evaluator must know what technological components are available, what capabilities they possess, and which components will be key in determining the success or failure of a proposed training device.

Both individual and collective training devices are complex systems consisting of many components that provide visual, auditory, and other sensory inputs; simulate the operation of a weapon system in response to operator controls and environmental conditions; and model the actions of other individuals and units. A detailed analysis of all potential technological components would likely be infeasible, and would almost certainly require greater effort than TPS Code analysis, which we have already criticized in this regard. Consequently, the method developed in this project must focus on the subset of components that are most likely to indicate the success or failure of the technology to meet training needs.

The components that are not considered by the method are not unimportant and may require considerable analysis when a training system is actually designed. For example, determining the requirements for operator controls and displays requires considerable effort and analysis. Some controls or displays need to be exact physical replicas of the actual equipment, because any departure from this high level of fidelity might lead to negative transfer of training. Others need to provide the same information or perform the same action as actual equipment, although they may be different physically. For example, a mechanical switch may be represented in a training system by a switch simulated on a touch screen. In addition, some controls may not be used to perform the tasks that are being trained, but provide tactile cues used by the operator to locate other controls or displays that are used. Finally, some controls or displays do not need to be represented at all, or need only to be represented by a drawing or picture. The required level of fidelity may need to be assessed individually for each control or display included in the operator workstation or cockpit. Such an analysis was performed to support the development of the AVCATT-A System Requirement Document (SRD; Simulation, Training and Instrumentation Command [STRICOM], 1999)

The previous discussion illustrates the complexity of the analysis that must take place to determine the appropriate level of fidelity for an operator workstation or cockpit. However, even though the overall operator workstation fidelity is important for determining the overall training effectiveness of the training system, it does not limit whether tasks can be trained using virtual environment technology. The ability to represent the workstation is sufficiently great to replicate nearly all displays and controls with nearly complete fidelity. Consequently, this component would not prevent a set of tasks from being trained using a virtual M&S. The focus of the analysis methodology developed in this effort is on less mature technologies that may not be sufficient to meet specified training requirements.

Restricting the method to a limited number of system components signifies one difference between a methodology that is designed to evaluate an existing or proposed system from a methodology that is designed to evaluate the general capabilities of virtual environment technology. For example, SIMNET training devices use a very austere representation of vehicle crew stations. The crew station design of SIMNET substantially limits the kinds of tasks that can be performed or trained using the system. However, even at the time that SIMNET was developed, the capability for high-fidelity crew station design existed and could have been used to produce a much more capable training system, albeit at a much higher cost. Thus, an evaluation of the capabilities of SIMNET would show deficiencies that would not be indicated by an overall evaluation of virtual environment technology to train the same tasks.

The specific index that is used to measure the capability of a critical component can have a substantial impact on how accurately the measure reflects the capability to perform or train tasks in a virtual environment. For example, a particular activity might require both a high visual resolution and a large visual field of view. This combination of requirements may put it beyond the capabilities of some virtual environment technology. However, the activity might require the highest level of resolution only in the center of the field of view, and thus be within the capabilities of a visual display system with a variable level of resolution. A measurement of visual resolution that did not distinguish central and peripheral resolution would not adequately characterize the capability of the technology with respect to the requirements of the activity.

Our effort to identify critical technology components of virtual environment training systems, determine appropriate performance measures, and characterize component performance according to these measures was based on three information sources. First, we surveyed selected training systems located at Ft. Knox, KY and Ft. Rucker, AL. The goal of this survey was to obtain general information about system capabilities, knowledge about how each surveyed system was used, and opinions regarding needs for improvements. The details of this survey are presented in Appendix A. Second, we reviewed earlier studies that identified technological components of virtual environment training systems, including two evaluations that identified problems in existing simulations. These analyses (Burnside, 1990; SHERIKON, 1995) determined the extent to which specific task activities could be performed or trained on SIMNET and CCTT, respectively. More important, the analyses identified specific reasons that activities could not be performed on the respective training systems. Finally, we obtained an overview of training system component capabilities from *Jane's Simulation and Training Systems* (Strachan, 1998, 1999). This source includes numerical summaries of the capabilities of a wide variety of training systems and components.

Characteristics of Virtual Environment Training Devices

To characterize the capabilities of virtual environment technology, we first identify the relevant attributes that describe simulation capabilities. Then we review the level of performance that is possible for each attribute.

Attributes of Simulation Capabilities

Simulation capabilities are often described in terms of major simulator components required to produce a virtual environment to be used for training, or in terms of the sensory cues and response feedback that the system must provide to meet training requirements. At the most general level, there is relatively close agreement between these two descriptions. For example, one general summary of simulation and training systems (Strachan, 1998, 1999) considers the following major simulation components: image generation system, visual display system, motion cueing system, and other simulator components. The list of cue and response attributes developed for the OSBATS model (Sticha et al., 1990), shown previously in Table 4, indicates a reasonable level of correspondence between the two taxonomies. This similarity suggests that these general categories may provide a useful starting point for the definition of the attributes that define simulation capability.

The specific attributes shown in Table 4 were developed primarily to characterize the capability of the flight simulators that existed at the time the OSBATS model was developed. Thus, these attributes may not address the new applications of virtual environment technology for training, including networked simulators and training of dismounted soldiers. Furthermore, the range of values does not reflect the substantial technological advancement that has occurred in the past decade. The capabilities of existing technologies are discussed in a later section. The following discussion identifies additional technology components from other sources.

Components for networked simulation. Because the individual simulators that are components of networked simulators are essentially the same as individual training devices (although they may not have all of those systems' capabilities), many of the attributes of individual training devices are relevant to networked simulators. However, the advent of networked simulators has introduced new attributes that must be considered in system design, including the following:

- Communication,
- Computer-generated forces (CGF) or semi-automated forces (SAF),
- Operator/controller (O/C) stations, and
- After-action review (AAR).

Some of the relevant issues related to these attributes are provided in the following discussion.

The communication capability of SIMNET represented a significant reduction from the state of the art at that time. This reduction was reasonable, given uncertainty about the feasibility of the technology and the need to minimize costs. Communication within the unit was implemented using CB radios. This solution is unrealistic in several respects that reflect potential attributes for describing communication capabilities. First, the equipment did not resemble the actual equipment in appearance or operation. Second, the quality of communication did not respond realistically to distance, environmental conditions, and opposing force activities. Third, communication networks often did not realistically represent the networks that a unit would use. The effects of unrealistic networking are that some participants in a SIMNET exercise may not have access to communication traffic that they would receive in a live exercise or in actual combat.

More recent networked simulators have provided more realistic simulations of unit communication by replicating the Single Channel Ground/Airborne Radio System (SINCGARS) unit in each simulator, and by using a hard-wired network to handle communications. The system can simulate the degradation in communication due to distance, terrain, and the effects of battle damage. Additional factors that affect communication, such as atmospheric conditions and electronic countermeasures (ECM) might also need to be simulated to adequately train some collective tasks.

The previous discussion suggests several attributes of the communication system that might be relevant in determining the ability of virtual environment technology to provide training for collective tasks.

- Fidelity of the communication equipment,
- Number and configuration of communication nets,
- Realistic degradation due to environmental factors, behavior of opposing forces (OPFOR) or equipment malfunctions.

Whether any or all of these components should be included in the evaluation methodology depends on the likelihood that they might present a barrier to virtual environment training for a given set of training requirements.

CGF are used to simulate the actions of opposing force units, as well as elements of friendly units that are not represented by actual soldiers. They may be used in individual or crew trainers as well as in collective training systems. For example, the gunnery targets that are presented to tank crews in the Unit Conduct of Fire Trainer (UCOFT) or the Advanced Gunnery Training System (AGTS) may be considered a simple type of CGF. However, some type of simulated force is essential for most unit training in a virtual environment for several reasons:

- The unit being trained may not be able to fill all functions required to perform the training mission;

- The cost required to assemble enough individuals to fill OPFOR rolls may be excessive;

- The virtual environment training system may not include person-in-the-loop simulation capability for certain functions; and

- There may be an insufficient number of available workstations to represent all participants in the training mission.

The CGF requirements for effective training have not been formalized. The *Close Combat Tactical Trainer (CCTT) Accreditation Report* (1999) indicates that the capabilities of the CGF for that system are acceptable for its training requirements, although it indicates potential for risk in several areas related to target acquisition, rate of aimed fire, maximum weapon range, and vulnerability to indirect fire. However, the Operational Requirements Document (ORD) for the One Semi-Automated Forces (OneSAF) program (STRICOM, 2000) indicates several problems with current CGF, including the CCTT SAF. Some of these problems were related to issues of interoperability, ease of use, and compliance with standards. Other concerns were related to fidelity, flexibility, and validation of automated behaviors. These later concerns might have an impact on the ability of the technology to meet training requirements, and consequently may be useful in defining a methodology that assesses the ability of virtual environment technology to meet training requirements.

The OneSAF ORD characterizes system requirements in several different ways, including (a) forces represented, (b) range of military operations, and (c) core physical models. The forces represented describes the types and levels of units that the OneSAF should be able to model for both friendly and opposing forces. The description of required forces specifies what types of units should be represented and the level of fidelity to which they should be represented. The ORD specifies that the CGF should be able to simulate friendly force operations from individual to battalion level for a number of combat, combat support[4], and combat service support units. The minimum acceptable fidelity of the simulation is specified by Mission Training Plans (MTPs), while the fidelity objective is based on Tactics, Techniques, and Procedures (TTP) for the unit. Opposing force individuals and units up to the brigade level must be modeled to represent a wide range of possible conflicts in a way that reflects authoritative sources describing opposing force operations and tactics.

A second way to categorize CGF capabilities is by the range of military operations that can be simulated. The OneSAF ORD is quite inclusive in the functions to be simulated but presents little detail in describing these functions. Combat functions and sub-functions included in the Army Universal Task List (AUTL) are to be simulated. In addition the simulation must represent information operations for both friendly and opposing forces.

The third characterization of CGF capabilities is related to core physical models. These models represent essential functions of combat that apply to a variety of units. Functions may be represented at different levels of aggregation (or fidelity) for units at different levels. The OneSAF ORD specifies the desire for the following core physical models.

- Target acquisition, including all sensors, radar, targeting devices, surveillance platforms, and identify friend or foe (IFF);

- Direct-fire delivery accuracy for predicted fire, guided, or smart weapons fired in a single shot or burst mode;

- Direct-fire vulnerability using standard vulnerability metrics for ground vehicles, rotary wing aircraft, and personnel;

- Indirect-fire vulnerability for ground vehicles , for both observer-adjusted and predicted fire using unguided projectiles;

- Indirect-fire delivery accuracy for predicted fire, guided, or smart weapons;

- Direct-fire rate of fire, considering crew proficiency, battle conditions, target range, time of fight, target motion, and differences between first round and subsequent rounds;

- Indirect-fire rate of fire, considering communication time between the firing unit and the Fire Direction Center (FDC), processing time as a function of unit and mission type, load and reload times, target range, and time of flight;

---

[4] Up to the Company level.

- Reliability, including firepower, sensor, electrical, and mobility failures;
- Mobility/countermobility based on the North Atlantic Treaty Operation (NATO) Reference Mobility Model;
- Line-of-sight (LOS) based on terrain and target posture;
- Communications;
- Combat Service Support (CSS), including maintenance, transportation, supply, ammunition, liquid logistics, medical, host nation and non-governmental support;
- Countermeasures (CM) and Counter-countermeasures (CCM) affecting communications, target acquisition, information flow, weapon system delivery to the target, and operations;
- Command, control, communications, computers, and intelligence (C4I) systems and structures; and
- Hazard and environmental sensors including Nuclear, Biological, and Chemical (NBC) sensors, infrared (IR) detectors, and meteorological sensors.

The best way to characterize the capability of CGF used as a component of virtual environment training systems is not clear. The three approaches described in the previous discussion do not appear to represent the issues that are of primary importance for determining training effectiveness. However, the OneSAF is being designed to serve several purposes and will form the basis of constructive simulations used for analysis and acquisition, as well as for virtual training simulations. It could be argued that these other applications place more stringent requirements on the capabilities of the CGF than virtual environment training does, because virtual environment training uses more human intervention into the activities of the CGF. In addition, training is often best conducted using structured scenarios that limit the range of actions available to the CGF. Thus, the most difficult problem for a CGF, simulating the commander's decision-making process, can be assigned to the operator or the training scenario developer. Consequently, it seems reasonable to expect that a CGF that is sufficient for constructive simulation and analyses will have capabilities for reasonable effectiveness in a virtual environment training system.

Despite this expectation, there are many ways that low fidelity CGF can hinder training effectiveness. For example, a mismatch between CGF weapon firing range and visual system range can greatly disturb the likelihood of a fair fight between human and computer-generated forces. Other factors, such as inappropriate movement techniques or rules of engagement can decrease training and transfer effectiveness.

Since collective training focuses skills in a different context than individual training, different types of instructor support features may be useful. The task simplification strategies that are appropriate for difficult individual psychomotor tasks, such as a variety of freeze capabilities, are not as useful in the collective environment where the primary focus is on cognitive tasks. For collective tasks, replay is a key instructional feature. It is very nearly always used in collective simulations, such as SIMNET and CCTT, while it much less likely to be used in individual trainers, such as the M1A1 driver trainer. Thus, it seems likely that the instructional support features that are selected for collective trainers will be different from those that are most useful for individual trainers.

Components for dismounted soldiers. Many unit missions involve activities conducted by dismounted soldiers. The design of CCTT has accommodated the need for dismounted soldiers by the development of a dismounted soldier station. The occupant of this station uses a joystick and other controls to move on the terrain, employ weapons, and communicate with others. Although this level of simulation may be appropriate for many collective training tasks where individual dismounted tasks are already learned, a higher level of fidelity would be required to train these individual tasks.

Jacobs et al. (1994) examined 292 unique individual activities contained in ARTEP mission training plans and drills for infantry and special forces units. The goals of their analysis were to determine the technological capabilities required to train these activities in a virtual environment and to estimate the effort required to develop the required capabilities. One result of their analysis was a description of current and future technological capabilities in several areas, as summarized in Table 5. As the table illustrates, some of the characteristics that Jacobs et al. identified, particularly those related to visual and auditory cues, correspond to characteristics of weapon system training devices listed in Table 4. However, other cues – particularly tactile, force feedback, or olfactory cues – apply primarily to dismounted soldiers (or apply to dismounted soldiers differently than they do to soldiers in a vehicle).

Sticha, Campbell, and Schwalm (1996) examined virtual environment interface requirements for combat leader training with a focus on speech recognition, gesture recognition, and CGF. Speech recognition capabilities were characterized by the following four factors: (a) trained vs. speaker-independent recognition, (b) isolated words vs. continuous speech, (c) vocabulary size, and (d) noise tolerance. This description of capabilities is similar to the description by Jacobs et al. (1994) shown in Table 5. Gesture recognition was characterized by the technologies used to detect the position of the trainees' hands and arms, and the processes that are used to recognize gestures once the position had been determined. Finally, the CGF features considered were specifically concerned with training dismounted combat leaders and included the following items: (a) information processing, (b) gesture recognition, and (c) human representation. The authors concluded that the capabilities available at that time were sufficient to perform some basic and intermediate scenarios, but that more advanced scenarios would require additional developments in the technology.

Summary. Previous research and analysis has identified several characteristics of simulation environments that can be used to define their capability. These characteristics relate primarily to the presentation of sensory cues. The range of performance levels considered in the OSBATS model do not consider many technological improvements that have occurred since the development of that model. However, the characteristics themselves may still represent useful considerations, with an appropriate updating of performance levels. Additional considerations that are relevant to collective training include communication networks and CGF, although it is unclear what features of CGF should be considered. In general, the requirements for dismounted personnel are more stringent than those for personnel in vehicles.

Table 5
Virtual Environment Technology Capabilities (from Jacobs et al., 1994)

Simulation Attribute	Range of Values	
	Low	High
Vision		
Type of Visual Display	Monocular	Binocular
Field of View	20°	120°
Resolution	300H x 200V pixels	400V x 3000H pixels
Scene Complexity	1,000 polygons	500,000+ polygons
Acoustic		
Speech Recognition	50 utterances, speaker dependent	5000 vocabulary, speaker independent, connected speech
Number of Sound Channels	1 channel	100 channels
3-dimensional Sound	None	Individual head-related transfer function, including echo, ambiance
Tactile		
Tactile Cues	Single Bladder for hand	Variable resolution, at least 200 x 200 elements for fingers and hand
Force Feedback		
Force Feedback	None	Bodysuit with active viscosity materials
Olfactory		
Olfactory Cues	None	Real time chemical synthesis of 100 odors

Critical Components Identified by Previous Analyses

Two analyses of the capabilities of SIMNET and CCTT have identified the areas that were judged by SMEs to represent deficiencies of the technology. Meliza (1993) summarized an earlier evaluation by Burnside (1990) of SIMNET to identify perceived deficiencies of the system that had a significant effect on training effectiveness. The top part of Table 6 shows the six problems that affected the rated effectiveness of the most tasks. The results of a similar evaluation of CCTT by SHERIKON (1995) are shown in the bottom part of the table.

Two things should be noted about these problem areas. First, the identified problems are not closely related to the sensory dimensions that were proposed by Sticha et al. (1990) and Jacobs et al. (1994). There are no comments about inadequacies of the visual system, audio special effects, motion cues, or other sensory factors. This result is especially surprising for

SIMNET, which has very significant limitations in its visual representation of the virtual battlefield and in the range of controls available in individual vehicle simulators. These capabilities were improved greatly in CCTT, but the improvements do not seem to be reflected in differences in the evaluations of the two systems.

Table 6
Top Problems Identified in SIMNET and CCTT Analyses

Problem Identified	Number of Tasks Affected
SIMNET (Meliza, 1993)	
Dismounted Personnel	32
Mines/Obstacles	19
Mark Terrain	9
Machine Guns	8
Hand and Arm Signals	5
Turret/Hull Down Positions	5
CCTT (SHERIKON, 1995)	
Dismounted Crewmembers	12
Nuclear, Biological, & Chemical Environment	11
Manipulate Individual Combatants	10
Manipulate Equipment	8
Mark or Manipulate Terrain	7

There are several possible explanations for these results, and little information to distinguish them. However, conjectures can be made to reconcile the results of these evaluations with the sensory requirements that were developed in earlier research. One possibility is that the evaluation method encourages the evaluators to identify concrete capabilities that are missing, rather than the more general ability to present appropriate cues. In addition, although high levels of cues may be required for crewmembers to perform individual tasks, they might not be required to the same extent for the collective tasks that were rated in these two studies. Because the two studies evaluated collective tasks only, they presented an incomplete picture of the problems that may affect the overall training effectiveness of the system. For example, they do not document known problems of SIMNET in detecting targets (because of an arbitrary visual range limitation) or identifying them (because of overly simplistic vehicle representation), because identifying targets is an individual task that was not rated in the evaluation. Similarly, the ratings of CCTT do not reflect the affects of the substantial improvement in those areas.

It is axiomatic that the focus of collective training systems is on collective tasks. However, it is also true that collective tasks are accomplished primarily through the coordinated efforts of soldiers doing their individual jobs. In many cases, the soldiers already know their jobs, and the simulation environment need only provide sufficient fidelity to avoid negative transfer of individual skills. However, the appropriate tradeoff between individual and collective requirements, and the levels of technical sophistication required to support individual tasks in a collective training environment have seen little research.

We think that both collective and relevant individual tasks should be considered when the capabilities of virtual environment technology are compared to training requirements. We anticipate that different factors will be relevant to collective tasks than are relevant to individual tasks. The appropriate level of cues for individual tasks conducted in a collective environment can not be estimated based on empirical performance, and will need to be estimated based on the experience of relevant experts.

A second characteristic of the results presented in Table 6 is the similarity of the two lists. Some of the most critical problems for SIMNET, such as the ability of crew members to dismount and to modify (e.g., camouflage) their equipment or the terrain is still a limitation of CCTT. These factors are obviously important considerations in evaluating virtual environment technology, because they represent limitations of the technology that currently have not been solved in production training systems.

Capabilities of Existing Technology

The levels of technology described in Table 4 and Table 5 represent an assessment of the capabilities of the technology available at that time and a prediction of the future course of technological development. Since the time of those studies, there have been significant advances in many of the areas included in those studies. Capabilities have increased in all areas. In some areas the capability has reached a sufficiently advanced level that the area is no longer an important consideration in training system design and evaluation. In the following discussion, we give a brief summary of the capabilities of some of the relevant technologies compared to the levels identified by Sticha et al. (1990) and Jacobs et al. (1994). Comparable performance data were not available for all display systems. Consequently, our summary focuses on selected systems from major manufacturers in which the reported performance data included variables that were important to assessing the capabilities of the technology for training simulations. Information regarding technical capabilities is from Strachan (1998, 1999).

Visual image generation. Substantial improvements in visual image generation systems have occurred during the last decade. Personal computer (PC) image generation systems in 1998 had the capability to generate 50,000 textured triangles at an update rate of 60 Hz. Dedicated image generation processors have even greater capabilities. This capability may approach the higher levels of scene complexity proposed by Jacobs et al. (1994).[5] They can apply large texture maps from satellite photography to terrain polygons to produce a highly realistic image. For example, one system introduced in 1996 can model up to 2,000 moving objects, each with up to 3 articulated parts. Other image generation systems have somewhat different capabilities. For example, a system developed in 1995 can model 255 objects, including 24 simultaneously moving models with 8 levels of articulation. This capability would be sufficient for vehicles, but would not be adequate for displaying dismounted personnel, because of the limit in the number of articulations.

---

[5] Although Jacobs et al. (1994) do not specify whether polygon count represents the number per frame or the number updated per second, we assume that this number is a per-frame value.

Regarding the specific variables developed by Sticha et al. (1990), typical performance of modern image generation systems meets or exceeds the highest levels addressed. The database size for image generation systems is limited only by the cost and effort required to obtain and process terrain information. Image generation systems routinely use both high-quality photographs and generic texture elements to represent terrain and cultural features. A wide variety of special effects are supported by these systems to represent environmental conditions and effects of the battle.

With the exception of the display of dismounted individuals, image generation capability does not appear to place any constraints that would affect training effectiveness of virtual environment training systems. The capabilities of existing video game systems to provide realistic representations of human motion suggests that this capability should be widely available in the near term, assuming that sufficient resources are dedicated to its development.

Visual display. A variety of projection displays, collimated displays, and helmet mounted displays (HMDs) can be used to provide the visual information required by a virtual environment simulation. The relevant variables for evaluating the visual display are the resolution, the FOV, and whether the display provides a binocular display (for HMDs only). Regarding resolution, some projection and HMDs allow a variable resolution, which is highest in the center of the FOV and lower in the periphery.

Projection display systems can offer both relatively high FOV with reasonable resolution in the center. For example, a display system developed in 1992 has a center resolution of 6.9 arcmin/OLP, while having peripheral resolution of 20.6 arcmin/OLP. That system has a 120° horizontal and 90° vertical FOV. A display from another manufacturer has somewhat better resolution and a larger FOV. At 2 arcmin/OLP, the resolution is near the maximum foveal acuity of the eye, which equates to approximately 1 arcmin/OLP. The FOV for this system is 270° horizontal and 130° vertical.

HMDs offer similar FOVs and may provide better resolution than projection systems. For example, one fiber optic HMD offers a 125° horizontal by 67° vertical instantaneous FOV with 38° overlap between the images presented to the two eyes allowing a stereoscopic view of objects in the overlapping region. This system has a central resolution of about 3 arcmin/OLP in an area-of-interest channel that is 24° horizontal by 18° vertical.

The FOVs of both projection displays and HMDs meet the requirements specified by both Sticha et al. (1990) and Jacobs et al. (1994). Consequently, the FOV is not a concern in evaluating the capabilities of virtual environment technology. Resolution seems adequate for all but the most demanding tasks. Evaluation methods need only identify those tasks requiring especially fine visual judgments.

Motion cueing systems. Motion cueing systems have not seen the dramatic progress over the last decade that has occurred with electronic systems. However, recent advances in the technology have included the use of electric rather than hydraulic actuators, and the reduction in response latency to the 10-13 msec range. Recent simulations that have used platform motion have represented all six degrees of freedom (DoF; pitch, roll, yaw, heave, sway, and surge). In addition to platform motion, there are several devices that provide seat motion, including full 6-

DoF seat motion systems. Such systems may be appropriate when motion cues are required, but there are space constraints that do not allow the use of platform motion.

The major uncertainty regarding motion cues regards whether they are needed. There has been considerable controversy regarding the utility of platform motion (see Boldovici, 1993). There seems to be a consensus that motion cueing devices that are poorly implemented with excessive latencies are probably worse than no motion at all. In addition, acceleration cues are the primary or earliest indicator of some malfunctions in aircraft and ground vehicles. Thus, although the decision to incorporate motion cueing in a training system is an important determiner of the cost-effectiveness of the system, motion may not be a factor that limits the capability of virtual environments to meet training needs.

Audio effects. Current sound sampling techniques combined with processing power supports the presentation of a variety of continuous and discrete sound effects. Current capabilities can process multiple moving sound sources and generate three-dimensional sound through speakers or headphones. Consequently, the requirements of simulation of vehicles are met by current technology, although further development appears to be required to provide individual three-dimensional direction information to dismounted personnel.

Speech recognition. The capability of speech recognition continues to improve as software becomes more robust. Commercial speech recognition software designed for PC applications can reach a word error rate approaching 5% for continuous speech in a trained situation. Speaker independent speech recognition has a somewhat higher error rate. For example in a 1998 benchmark study, Pallett, Piscus, Garofolo, Martin, and Przybocki (1999) found that the best system had a word error rate of 9.7% in recognizing speech from broadcast news reports in baseline conditions. The error rate increased to 14.4% under degraded acoustic conditions, and doubled to 19.5% when music was played with the speech. Furthermore, the processing time required to obtain this error rate was ten times the actual time of the speech sample. Consequently, it seems that for the moment, real-time speech recognition is appropriate for simple commands and phrases only.

Tactile and force cues. Jacobs et al. (1994) focus on tactile cues presented to the hand. Existing technology provides some capability in this area, but further analysis would be required to determine whether current capabilities are sufficient to simulate activities performed by dismounted soldiers. Instrumented gloves can sense the position of the fingers and, optionally, can provide tactile feedback using a vibrator attached to the palm, thumb, and each finger. In addition, force feedback can be provided using an exoskeleton that provides pressure to the hand and fingers. Clearly, tactile and force cues might be a limiting factor for some tasks performed by dismounted soldiers.

Characterization of Technology Capabilities and Training Requirements

Research over the past 10 years has identified about 30 characteristics that might be used to define virtual environment capability. However, technological capabilities have advanced to the point in which many of these characteristics are no longer barriers to the implementation of virtual environment training systems. The following list summarizes the factors that still present a potential difficulty for some problems.

- Resolution of the visual display;
- Scene complexity, particularly for dismounted soldiers in an urban environment;
- Automated speech recognition;
- Automated gesture recognition;
- Tactile and force cues for dismounted individuals;
- Ability of crewmembers to dismount;
- Ability to manipulate terrain; and
- Ability to manipulate equipment.

Not all technology dimensions are relevant to all training domains. The dimensions on the list are weighted towards factors that are critical for representing dismounted soldiers. The technology used to represent soldiers in vehicles is more mature, and many of the relevant components have been developed to the extent that they no longer provide a meaningful limit to the representation of the tasks in the training domain.

SATISFYING TRAINING REQUIREMENTS IN VIRTUAL ENVIRONMENTS (STRIVE)

The goal of the method for Satisfying Training Requirements in Virtual Environments (STRIVE) is to estimate whether the capabilities of virtual environment technology are sufficient to allow training of a specified set of activities. STRIVE modifies and extends TPS code analysis by incorporating a simple behavioral analysis that is focused on the aspects of military tasks that are likely to present difficulty to virtual environment training.

Like the TPS code, STRIVE estimates task performance support based on SME judgments. However, STRIVE incorporates several changes in both the nature and number of the judgments that are required to calculate the estimate of support. These changes are made to allow a more detailed behavioral description of activities to be assessed without dramatically increasing the number of judgments that are required. Rather than making direct judgments about whether a given technology can support the performance of a specified task element, the rater must answer several questions regarding the type of activities that are required. The particular behaviors assessed include taxing visual tasks, such as detecting small or distant objects or making visual distance estimates, communicating through gestures or hand and arm signals, or making modifications to terrain or equipment. Other questions may assess the sensory cues and feedback required to perform a particular task or task element.

The method has been designed to minimize SME effort in several ways. First, STRIVE allows raters to assess more aggregated activities, including tasks and task steps, instead of restricting them to rate individual performance measures. Second, the technology dimensions considered in the method are restricted to those that are likely to present a problem to virtual environment training. The number of technology dimensions is further restricted by the training domain. For example, certain technology dimensions are only considered when the training domain includes tasks that are performed by dismounted individuals interacting with terrain.

This section presents both an overview of the STRIVE method and a more detailed description of the steps in the process. It also describes the two example problems that were developed to illustrate the procedure.

Overview

The Integrated Definition (IDEF0) system analysis procedure was used to illustrate the major components of the STRIVE method (see Figure 1). The IDEF0 procedure breaks a complex system into activities – which are represented by boxes – and the input, controls, output, and mechanisms associated with these activities – represented as arrows that connect boxes. A complete IDEF0 breaks a complex system down hierarchically using a series of diagrams with an increasing level of detail. For the purpose of this overview, Figure 1 presents the first level of decomposition of the method that illustrates the five basic steps in the procedure. In describing each step, we provide a general discussion of what the activity produces, what it uses for input, and how it is constrained by other information.

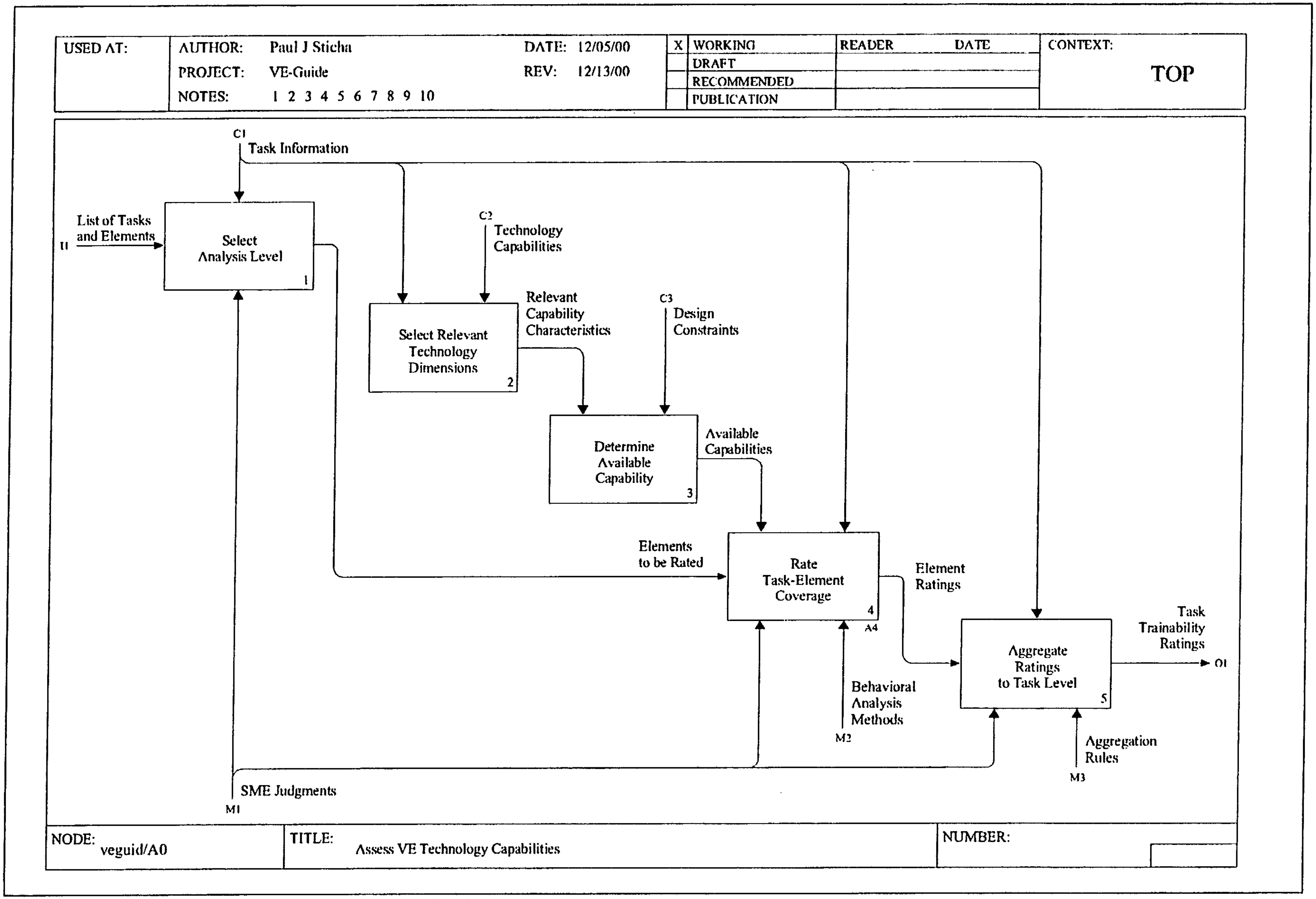


Figure 1. IDEF0 description of STRIVE method.

Select Analysis Level

This activity determines whether the elements rated in the analysis should be tasks, task steps, performance measures, or some combination of these three. The input data for this process are a list of tasks, each of which is broken into task steps and performance measures (this terminology is taken from ARTEP MTPs; several other terms could also be used). The SME decides, based on task documentation, which tasks can be rated as single units, which tasks can be rated at the task step level, and which steps need to be broken down further to individual performance measures. The output of this step is a list that includes the tasks, task steps, and performance measures that will be rated.

Select Relevant Technology Dimensions

Of the eight dimensions that were selected to describe capabilities of virtual environment technology, three are appropriate only when dismounted activities are required, and one is appropriate only when members of vehicle crews must dismount. The task information that controls which technologies are selected simply assesses whether each of these two types of activities is required. If it is, then the appropriate technology dimensions are included, while if it isn't, those dimensions are excluded from the analysis.

Determine Available Capability

Certain design constraints limit the capabilities that are possible using some virtual environment technologies. For example the requirement for a simulator to be transportable limits its size and precludes the use of particularly large components, such as large projection display systems. Consequently, the capability of a transportable system will be restricted to the extent that capable alternatives are excluded. Other design decisions introduce technology dimensions into consideration. For example, the design may reflect a willingness to consider automated speech recognition to meet some of the communication needs of the tasks within the training domain. This step in the analysis evaluates those constraints and adjusts technological capabilities to reflect the effects of these design constraints.

Rate Task-element Coverage

In this step, the SME makes judgments about each of the selected task elements (which may be tasks, task steps, or performance measures). The behavioral analysis method that is the mechanism for this step consists of a set of questions to be answered for each of the technology dimensions. The available technology capabilities restrict the process by limiting the questions that are asked to those that relate to the technology dimensions included in the analysis. The answers to the questions imply a required level of performance for the technology dimensions. Based on these answers and the available capability for that dimension, a rating of performance support is calculated. Support is rated on a four-point scale with the following levels: no support (0), low support (1), moderate support (2), and high support (3).

Aggregate Ratings to Task Level

Because ratings are made at the task, task step, and performance measure level, the ratings must be propagated to both higher and lower levels of aggregation. Ratings made at the task level are simply duplicated for all steps and performance measures within that task. Similarly, ratings made at the task step level are duplicated for all performance measures within that step. Ratings made at the performance measure and task step levels are aggregated to the task step and task levels, respectively, using the methods derived from the CCTT Accreditation Report (1999). The SME provides importance weights that are used in the calculation of task support scores. The resulting scores can be compared at any level of detail.

Detailed Description of Analysis

This section presents a more detailed description of the analysis, in terms of both the questions that the SME performing the analysis must answer and the calculations that are used to obtain overall support scores based on those answers. Certain activities must be conducted before the method is applied. For example, tasks for training must be selected, and task documentation, such as ARTEP MTPs, must be obtained. In addition, the training population should be defined in order to determine which individuals will be trained, which will be present but serve primarily as training aids, and which will be represented by controllers or computer-generated forces (CGF). Any design guidance or other constraints should be collected to incorporate into the analysis, as appropriate. When the required information is available, the analysis may proceed with the following steps.

Select Analysis Level

The SME reviews the steps within a task and decides whether the activities are sufficiently similar that the task can be rated as a unit. If a task can be rated as a whole, then this activity is complete for that task, and the SME should proceed to the next task. If the steps are different, then the SME must continue by deciding whether ratings should be at the step or performance measure. For each step, the SME reviews the performance measures within that step and decides whether the performance measures are sufficiently similar so that the step can be rated as a unit. If the task step cannot be rated as a whole, then each of the performance measures in that step must be rated.

There are several reasons that the SME may decide to rate at a more aggregated level, rather than to rate more detailed subelements.

- The subelements are all highly similar regarding the type of activity they require or the kind of stress they place on virtual environment technology. For example, one potential task might be "Identify major components, controls, instruments, and indicators." The steps of this task specify the individual components that must be identified. Since the requirements to make these identifications do not vary appreciably with the component being identified, the task may be rated as a whole with little loss of accuracy.

- The SME understands the aggregated element very well and can easily rate it as a whole.

- The task places minimal requirements on virtual environment technology and can easily be supported. For example, simple procedural tasks may place little demand on the simulation technology. Because the entire task is obviously within the capability of the technology, making an assessment at the task level will be more efficient than making judgments at a greater level of detail.

- It is desired that the total number of judgments required be minimized. For example, an analyst may decide to perform a quick, preliminary analysis at the task level, even though the accuracy of this analysis would be reduced, compared to an analysis of more detailed task elements.

On the other hand, the SME may decide to rate more detailed subelements for one of the following reasons.

- More detailed elements differ significantly regarding the extent to which they can be supported by virtual environment technology. For example, driving a vehicle at night involves task steps that range from planning the route, which can be done without technological support, to applying specific night driving techniques, which may present greater challenges to virtual environment technology.

- More detailed elements differ in the kind of activity they require, making it difficult to rate as an aggregated unit. For example, driving a vehicle off road involves different procedures to negotiate streams, ditches, sand, mud, or rocky terrain.

The result of the step is a list of the items that will be rated. In general, the list will include tasks, task steps, and performance measures. However, in any specific application of the method, the list might include tasks only, task steps only, or performance measures only.

Select Relevant Technology Dimensions

Our review of technology dimensions indicated that some need be considered only when dismounted soldiers are being trained, while others are more generally applicable. Consequently, the STRIVE method asks some general questions about the training domain, and uses the answers to these questions to select the technology dimensions that will be evaluated in the following steps. In the remainder of this description, questions that are asked of the SME are shown in italicized text, with response options represented by bullets following the question. To determine the technology dimensions that will be considered, the SME is asked the following two questions:

> *Do the training requirements include tasks that require training participants to move or otherwise interact with terrain outside of a vehicle (i.e., dismounted)?*
>
> - *Yes*
> - *No*
>
> *Do any individual training participants need to perform some activities on a vehicle and other activities dismounted?*
>
> - *Yes*
> - *No*

The second question is only asked if the answer to the first question is "yes."

The following three technology dimensions are considered for all cases: (a) visual display resolution, (b) manipulation of terrain, and (c) manipulation of equipment. In addition, the following three dimensions are considered when the first question is answered affirmatively: (a) scene complexity, (b) gesture recognition, and (c) tactile/force cues. An additional dimension, ability to dismount, is considered in the analysis when the second question is answered affirmatively. Finally, speech recognition is determined in the next step.

Determine Available Capability

The purpose of this step is to determine whether there are any design constraints that might affect the performance that is possible using virtual environment technology. This information is assessed with the following questions.

Should automated speech recognition be considered to respond to communications between training participants and those not being trained?

- *Yes*
- *No*

Is the training system required to be transportable?

- *Yes*
- *No*

Must the training system be reconfigurable to represent different vehicles or different versions of the same vehicle?

- *Yes*
- *No*

The first question assesses whether automated speech recognition should be considered as an option to address communication needs. In many cases, automated speech recognition and the use of live controllers are alternative approaches to simulating communications between individuals in the training population and those who are not. It is beyond the scope of this method to determine which of these approaches should be chosen. Rather, the method relies on the SME to choose whether speech recognition should be considered. This choice can be based on design requirements, if they exist. If there are no requirements, then the choice reflects the preferences of the agency performing the analysis. If it is included, then the method will identify the tasks (and more detailed task elements) that require communication activities that may be beyond the current capability of automated speech recognition methods.

Transportability and reconfigurability both affect the allowable size of a training system, and may also affect its capability. For the purpose of the method, if either of these requirements exist, then the minimum resolution of the visual display system takes on the value of a helmet-mounted display, which is slightly less than that for a projection display. These requirements may also affect the capability for motion cueing, but that dimension was not included in the method.

If the STRIVE method is applied early in the design of a system, then the constraints may not be known. In this case, it is probably best for the SME to assume an unconstrained design that includes all technology dimensions that might be considered (i.e., speech recognition). The effects of adding constraints could then be investigated using additional analyses. Because the design constraints do not affect the requirements, their effects could be determined without performing additional ratings of the task elements.

Rate Task-Element Coverage

This step represents the bulk of the activity for the STRIVE method. It is in this step that the SME answers several questions regarding each of the task elements to be rated. The task elements are the tasks, task steps, or performance measures selected by the SME. Each question relates to one technology dimension. Based on the answers to the questions, the method calculates a score for each technology dimension that represents the extent to which the capabilities of that dimension support the activities conducted in the rated task element. The overall score for the task element is the minimum of the scores for the technology dimensions that are considered.

All calculated support scores are made on the following scale, which corresponds to the scale used in the CCTT Accreditation Report (1999, p. 49).

- 3 – High Support. Training is fully supported with physical cues and responses and/or does not detract from cognitive processes
- 2 – Moderate Support. Training is supported with physical cues and responses and/or minimally detracts from cognitive processes.
- 1 – Low Support. Training is marginally supported with physical cues and responses and/or may detract from cognitive processes.
- 0 – No Support. Training is not supported with physical cues and responses and/or detracts from cognitive processes.

Unlike the CCTT Accreditation Report, the STRIVE method uses the same numerical scale at all levels, because it allows SME ratings to be made at all levels. Consequently, tasks, task steps, and performance measures are rated on a common numerical scale.

We provide a description of the questions that address each technology dimension, and the calculations that are used to estimate support of the task element.

Visual display resolution. This evaluation considers the requirements for a high-resolution visual display. The questions address two activities that may require high resolution: detecting small or distant objects and estimating distances. The answer given to the following question determines which of these activities may apply to the task element being evaluated.

Please indicate whether the activity requires any training for participants to visually detect small or distant objects, or to make accurate visual estimations of distances.

- *No visually demanding activities*
- *Visually detecting small or distant objects*
- *Visually estimating the distance to an object*
- *Both visual detection and distance estimation*

If the SME selects the second or the fourth response option, then the following additional questions about the requirement for visual detection are asked.

Consider the most difficult visual detection that must be made for this activity (that is, the smallest object and/or greatest distance). Rate the minimum size of the object that must be detected and the maximum distance at which it must be detected.

Minimum size of object in meters: (numerical response)

Maximum Detection distance in meters: (numerical response)

If the SME selects the third or the fourth response option, then the following additional questions about the requirement for visual estimation of distances are asked.

Consider the most difficult distance estimation that must be made from this activity (the longest distance or lowest tolerance). Please rate the longest distance that must be estimated, the size of the objects used to estimate that distance, and the percentage tolerance allowed for that estimation.

Greatest estimation distance in meters: (numerical response)

Size of object to estimate distance to in meters: (numerical response)

Tolerance for error as a fraction of distance (0-1): (numerical response)

The answers to these questions are used to calculate a required display resolution in minutes of arc (arcmin). For detection, the calculations assume that the objects to be detected are small (otherwise there would be no problem with display resolution). In this case, the visual angle required to detect an object is the ratio of the size of the object to its distance.[6] The required resolution is limited to between 1 arcmin per optical line pair (OLP), which represents the limit of human foveal vision, and 12 arcmin/OLP, which represents a level of resolution that can be met by nearly any visual display system.

---

[6] This ratio estimates the resolution in radians. It must then be converted to the desired unit, such as minutes of arc.

Estimating distances or ranges can be a problem if the distances to be estimated are large, or if the tolerance for error is small. Visual angle is one of several factors that may be used to estimate distance. However, factors other than visual angle (e.g., texture gradients, binocular disparity) might also be used to estimate distance. Considering only visual angle provides the most accurate representation of display resolution requirements, when other factors are less important, such as when distances are relatively great or when visibility is relatively poor. In other situations, the equation used to calculate the resolution required to estimate distances will contain some error.

The STRIVE method estimates the required resolution using the following equation:

$$R = 2\left[\operatorname{atan}\left(\frac{s}{2d(1+t)}\right) - \operatorname{atan}\left(\frac{s}{2d}\right)\right],$$

where R is the required resolution, s is the size of the object, d is the distance to the object, and t is the tolerance expressed as a fraction. Similar to detection, the required resolution for estimating distances is limited to between 1 and 12 arcmin/OLP.

The overall resolution requirement is the minimum resolution required for detecting objects or estimating distances. Available systems have sufficient visual resolution to accommodate nearly all requirements. For transportable or reconfigurable systems, the method assumes the resolution of a helmet-mounted display, which is set at 3 arcmin/OLP, based on an available display from a major manufacturer. When there is no requirement for the system to be transportable or reconfigurable, the method uses an estimated resolution of 2 arcmin/OLP, based on an available high-resolution projection display system. If the resolution required by the task element being rated is greater than the available resolution considering design constraints (i.e., a requirement to be transportable or reconfigurable), then the task element is judged to be highly supported. Otherwise, the task element is moderately supported. Because of the high level of capability in this area, it is possible to support all requirements at least moderately.

Scene complexity. The number of polygons that can be displayed does not appear to represent a limit of visual image generation systems that affects training effectiveness. Consequently, the focus of this step is on the number of moving images that can be displayed simultaneously and the level of complexity of the objects that can be displayed (i.e., the levels of articulation). The SME would assess these requirements by answering the following questions regarding each rated task element:

What is the maximum number of independently moving objects simultaneously visible to any single individual performing this activity?

- *25 or fewer*
- *Between 26 and 200*
- *More than 200*

Is it necessary to display realistic movement of individual people (e.g., to show hand and arm signals or other gestures)?

- *Yes*
- *No*

Recent image generation systems can support up to 256 moving models, although older systems limit the number of simultaneously moving objects. However, these models typically have a relatively small number of levels of articulation (eight or fewer). That level of articulation can only represent a very simple representation of human motion. Consequently, the current generation of image generation systems can highly support task elements that require 200 or fewer simultaneously moving objects, if there is no requirement for realistic human movements. They can also provide moderate support to task elements that require more than 200 moving objects. This level of support could be provided by combining objects so that they would move together. Currently, image generation systems provide only low support for displaying realistic human motion. However, recent advances in computer gaming systems would suggest that a more advanced capability might be available in the near future.

Speech recognition. Because most requirements for automated speech recognition can also be handled by live controllers, this technology dimension is only considered when the design constraints call for it to be considered. Requirements for speech recognition are assessed with the following question.

> *Please indicate the extent to which individuals being trained need to speak to others who are not part of the population being trained (i.e., represented by CGF or controllers).*

- *None*
- *Isolated commands, vocabulary known in advance*
- *Commands or information embedded in continuous speech*
- *Unformatted messages in continuous speech*

In estimating the support provided in the area of speech recognition, the method assumes that the recognition system will be speaker independent, and will not be trained to individual characteristics. Given this assumption, only a need to recognize isolated commands is highly supported by the current capabilities of technology. The technology provides low support for understanding commands or information embedded in continuous speech and no support for understanding unformatted messages in continuous speech. However, like image generation, speech recognition is a technology that has seen substantial progress in the past few years, and is likely to see continued progress over the next few.

Gesture Recognition. This factor is appropriate for dismounted personnel (although it conceivably could be used in other situations). Requirements for gesture recognition are assessed using the following question, which was taken from an earlier description of capabilities in this area provided by Sticha, Campbell, and Schwalm (1996).

> *Are any individuals being trained required to communicate with others outside of the training population using hand and arm signals or other gestures? If so, are the gestures static or dynamic?*

- *No gestures required*
- *Static gestures only*
- *Dynamic gestures with or without static gestures*
- *Gestures correlated with voice*

Because gesture recognition is not an element of any operational training devices that we are aware of, the assessment of capabilities in this area is relatively uncertain. The method assumes moderate support for static gestures, low support for dynamic gestures, and no support for gestures correlated with voice.

Tactile/Force cues. This factor is appropriate for dismounted personnel only. It refers to tactile and force cues presented directly to the body, rather than the cues that are felt through controls. The need for tactile and force cues is assessed with the following two questions.

Please enter the maximum level of tactile cues required to train this activity.

- *None*
- *General cues to hands and fingers only*
- *Detailed cues to hands or fingers*
- *Tactile cues to body other than hand*

Please enter the maximum level of force cues required to train this activity.

- *None*
- *General pressure to hand or fingers*
- *Forces to other body parts*

Existing tactile and force cueing devices for the hand, such as instrumented gloves, can provide general tactile cues to the hands and fingers, but they have limited capability to provide detailed cues. Consequently, general tactile cues to the hand are currently supported, but support for detailed cues is low. Force cues to the hand are moderately supported by the types of vibrations that can be provided by current technology. Both tactile and force cues to other parts of the body are not supported. The overall level of support for this technology dimension is the minimum of the support for the tactile cues that are required and the support for the force cues that are required.

Ability to dismount. The requirement to dismount is assessed using the following question.

Do any individuals need to dismount or mount a vehicle or weapon system to perform this activity?

Currently, the ability of soldiers to be in a vehicle and subsequently dismount is not supported by virtual environment technology. Consequently, the method indicates that the task element is not supported by virtual environment technology when individuals are required to dismount.

Manipulation of terrain. The requirement for manipulation of terrain is assessed with the following question.

> *Does this activity require any individuals being trained to manipulate terrain (e.g., dig positions)? If so, must the individuals use their own equipment or vehicle?*
>
> - *Not required*
> - *Manipulation by other equipment of vehicle*
> - *Manipulation using own equipment or vehicle*

Current technology, such as the technology employed in CCTT, allows for some SAF vehicles to modify terrain – a level of support that it judged to be moderate. Manipulation using the trainee's own equipment or vehicle is not supported.

Manipulation of equipment. The requirement for manipulation of equipment is assessed with the following question.

> *Does this activity require any individuals being trained to manipulate their equipment (e.g., camouflage, or repair)?*

Currently this capability is not supported by virtual environment technology.

Determine overall support score. The answer to the preceding questions determines a performance support score for each technical dimension. The STRIVE method assumes that all technical requirements must be satisfied in order for performance of the rated task element to be supported by virtual environment technology. Consequently, the overall support score for a rated task element is the minimum score for the technology dimensions that were rated. This approach is consistent with the typical practice for TPS code analysis in which a single reason for failure to meet requirements is given.

Aggregating Ratings to Task Level

The fact that ratings are made at three levels of detail has several implications on the way that scores rated at one level are combined to produce a rating at a higher level. First, the measures must use the same scales so that directly rated scores at a given level may be compared to scores that are calculated from ratings made at another level. That it, the aggregation rules must produce scores for a task or step that are comparable to the scores obtained by direct ratings. Second, it is necessary both to aggregate scores to higher levels and to migrate them to lower levels. Consequently, scores assessed at the task or step level are duplicated at the step or performance measure level, respectively.

The three versions of the TPS code use slightly different methods to aggregate scores to higher levels. The version used by the STRIVE method is based on the procedures used by the CCTT Accreditation Report (1999, p. 51). However, STRIVE uses numeric scores throughout, rather than a combination of numeric and nominal scales. In addition, some changes were needed to ensure that direct and aggregated ratings were comparable.

Migrating ratings to more detailed levels. The fact that a rating is made at a general level (e.g., task level) generally indicates that all the subelements of the rated item have similar requirements. Consequently, scores of rated tasks are copied to all steps and performance measures within those tasks. Similarly, scores of rated steps are copied to all performance measures within those steps.

Aggregating performance measure scores to the step level. Following previous methods, the rules used to aggregate performance measure scores to the step level depend on the number of performance measures included in the step. The following rule is used when there are four or more performance measures. The conditions are evaluated in order. Thus, each step receives the highest score for which the relevant conditions are true for that step.

3	At least 66% of the associated performance measures must receive the rating of "3;" no performance measure receives the rating of 0 or 1.
2	At least 66% of the associated performance measures must receive the rating of either "2" or "3;" the remaining performance measures may have any rating.
1	At least 25% of the associated performance measures must be rated "1" or higher
0	None of the preceding conditions is met.

When there are three performance measures, the following aggregation rule is used.

3	At least two of the associated performance measures must receive the rating of "3;" the remaining performance measure must not receive the rating of "0" or "1."
2	At least two of the associated performance measures must receive the rating of either "2" or "3;" the remaining performance measure may have any rating.
1	At least one of the associated performance measures must be rated "1" or higher
0	None of the preceding conditions is met.

This rule is actually equivalent to the rule used when there are four or more performance measures in a step.

The following rule is used when there are one or two performance measures in a step.

3	The minimum rating of the associated performance measure or measures is "3."
2	The minimum rating of the associated performance measure or measures is "2."
1	The minimum rating of the associated performance measure or measures is "1."
0	None of the preceding conditions is met.

This rule combines two separate rules that were used by the CCTT Accreditation Report (1999). It also reflects one modification to the rule from that report for assigning a rating of "1" when there are two performance measures. Specifically, the CCTT Accreditation Report required at least one associated performance measure to receive a rating of "2" or "3," in addition to the other performance measures receiving a rating of "1." We rejected this rule because it did not assign a value of "1" to a step when both performance measures associated with the step received a rating of "1." That result was inconsistent with the corresponding rules when there were three or more performance measures, as well as with the desire for comparability in the meaning of the rating scale between the step and performance measure levels.

Assessing step weights. Step weights are used to aggregate task step scores to the task level. The SME is asked to indicate for each task step, the importance of training that step in a virtual environment. Ratings are made on the following four-point scale: Not important (0), low importance (1), medium importance (2), essential (3).

Aggregating step scores to the task level. In general, the task scores were derived from the task step scores using the following equation.

$$T_i = \frac{\sum_j w_{ij} \times s_{ij}}{\sum_j w_{ij}},$$

where T_i is the task score for task i, s_{ij} is the score for step j of task i, and w_{ij} is the importance weight of step j of task i. The scores obtained from this equation are rounded to the nearest integer to be comparable to directly assessed scores. Following the procedures of the CCTT Accreditation Report, there is one exception to this rule. Whenever a task has a essential task step (importance weight = 3) with a score of "0," that task also receives a score of "0."

The CCTT Accreditation report continues by converting the scores to a normalized scores and defining what is termed, bands of potential. In order to maintain a single scale for scores at all levels, the STRIVE method does not follow this part of the procedure.

METHOD DEMONSTRATION

A demonstration of the method was developed using Microsoft Access97. Two example problems were implemented in the model demonstration. The demonstration does not represent operational software and has several limitations in its use. Nevertheless, it serves to illustrate the method, and was used to obtain the assessments for the example problems. This section describes the capabilities and limitations of the STRIVE demonstration and summarizes its operations.

Overview of Capabilities

The STRIVE demonstration includes all the steps in the method described in the previous section. It allows the user to select the level at which to make ratings, specify the general requirements and design constraints, rate the selected task elements, rate task step importance, and calculate and display coverage scores. All of the calculations follow the specified procedures.

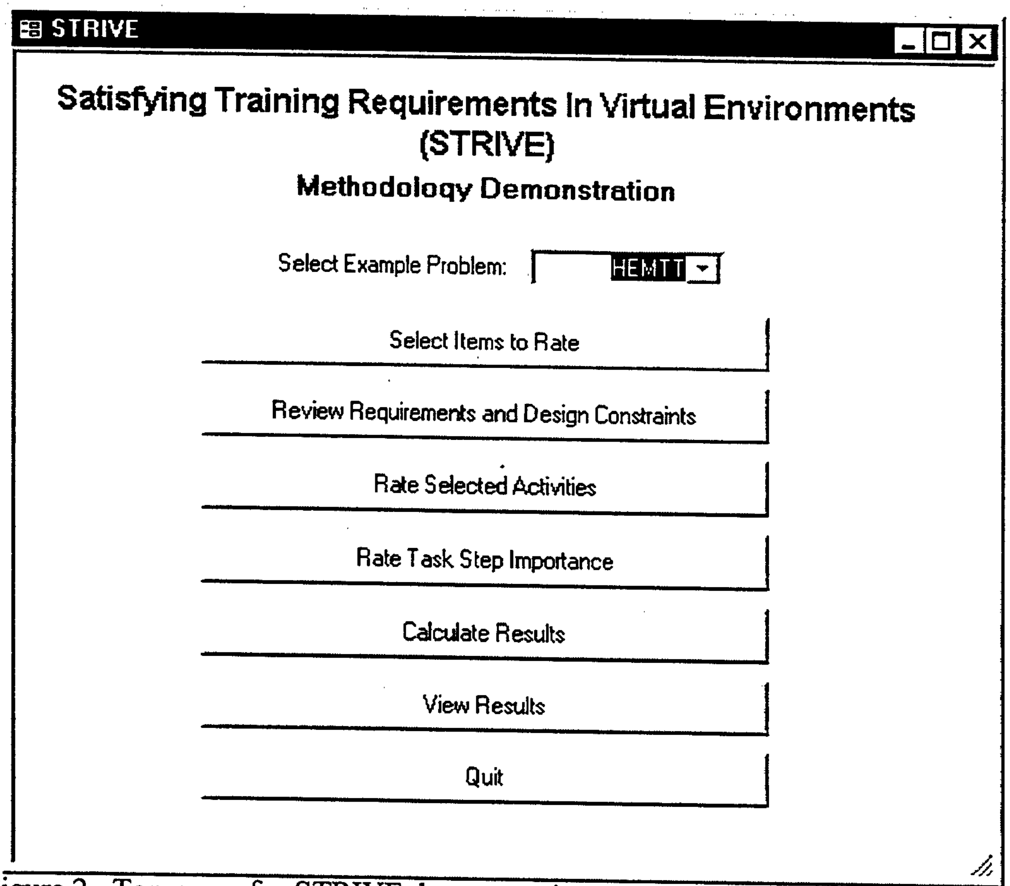
However, as a demonstration rather than operational software, there are several limits to its capabilities. First, certain functions that would be a part of operational software are not included in the demonstration. For example, there is no system for managing task data or for revising technology dimension capabilities. Second, the demonstration is designed to illustrate the steps of the method in a fixed order. Some deviations from this order may not produce the correct results or may erase rating data. Finally, the demonstration does not have the level of error checking and user support (such as provision of help) that would be available in operational software.

Summary of Operation

The STRIVE demonstration consists of seven activities that can be performed on two example problems. When the program is started, the main menu (Figure 2) lists the options that are available. The demonstration is designed to go through the options in order, which is how they will be described in this section.

The first step in the analysis is to select the example problem. Two example problems have been developed. The first includes nine tasks related to the AVCATT-A training system. Six of these tasks were selected from the AVCATT-A Operational Requirements Document (ORD). The other three were selected from the relevant ARTEP MTPs because they presented problems for virtual environment technology that were useful to illustrate in the demonstration. The second example includes eight tasks that are preformed by the operator of the HEMTT. There is no current ORD for a HEMTT training system. Consequently, these tasks were selected to represent a variety of situations to illustrate in the demonstration. The example problem is selected using the combo box located under the title of the main menu form.

Each of the buttons below the example problem begins one or more steps in the process. The following discussion describes these options in order.


STRIVE

Satisfying Training Requirements In Virtual Environments (STRIVE)

Methodology Demonstration

Select Example Problem: HEMTT

Figure 2. Top menu for STRIVE demonstration.

Select Items to Rate

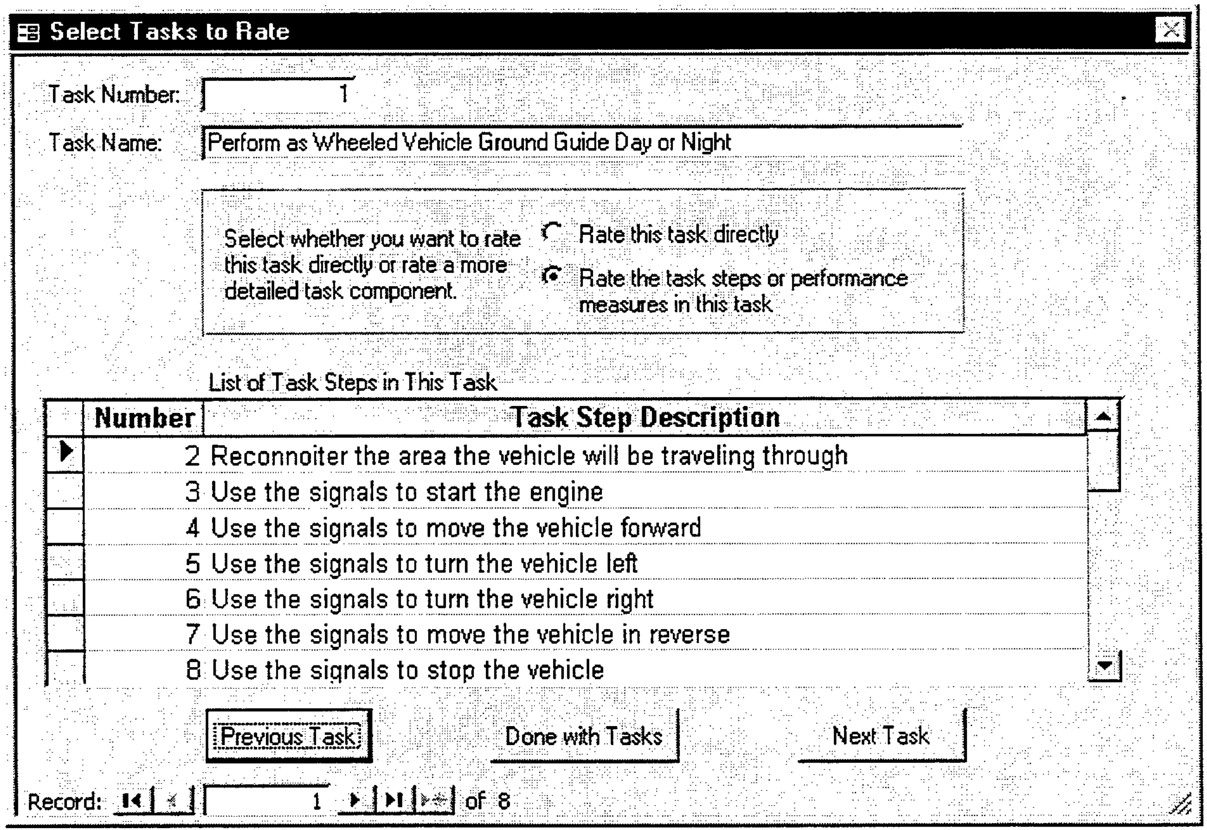
The first activity involves the selection of the appropriate items to rate, that is, tasks, task steps, or performance measures. The user first identifies the tasks to rate, using the form shown in Figure 3. This form shows the task name, and lists all of the task steps within each task. After reviewing the steps included within the task, the user answers the following question.

> *Select whether you want to rate this task directly or rate a more detailed task component.*

- *Rate this task directly.*
- *Rate the task steps or performance measures in this task.*

The user may decide to rate this task directly if the task steps are similar in their technology requirements, if the user understands the task as a unit very well, or if it is obvious to the user that the task is easily supported by virtual environment technology. Otherwise, the user should not rate the task, and should make ratings at the task step or performance measure level. For example, the steps in the task shown in Figure 3 indicate two types of activities:
(a) reconnoitering the area, and (b) giving a variety of hand and arm signals. Because these two types of task steps are different, the figure indicates that they should be rated separately.

After the user has made a selection on the first task, he or she should proceed to the next task until all tasks have been completed. When the task selections are complete, the user should press the button labeled "Done with Tasks" to select task steps.



Task Number: 1

Task Name: Perform as Wheeled Vehicle Ground Guide Day or Night

Select whether you want to rate this task directly or rate a more detailed task component.

○ Rate this task directly

● Rate the task steps or performance measures in this task

List of Task Steps in This Task

Number	Task Step Description
2	Reconnoiter the area the vehicle will be traveling through
3	Use the signals to start the engine
4	Use the signals to move the vehicle forward
5	Use the signals to turn the vehicle left
6	Use the signals to turn the vehicle right
7	Use the signals to move the vehicle in reverse
8	Use the signals to stop the vehicle

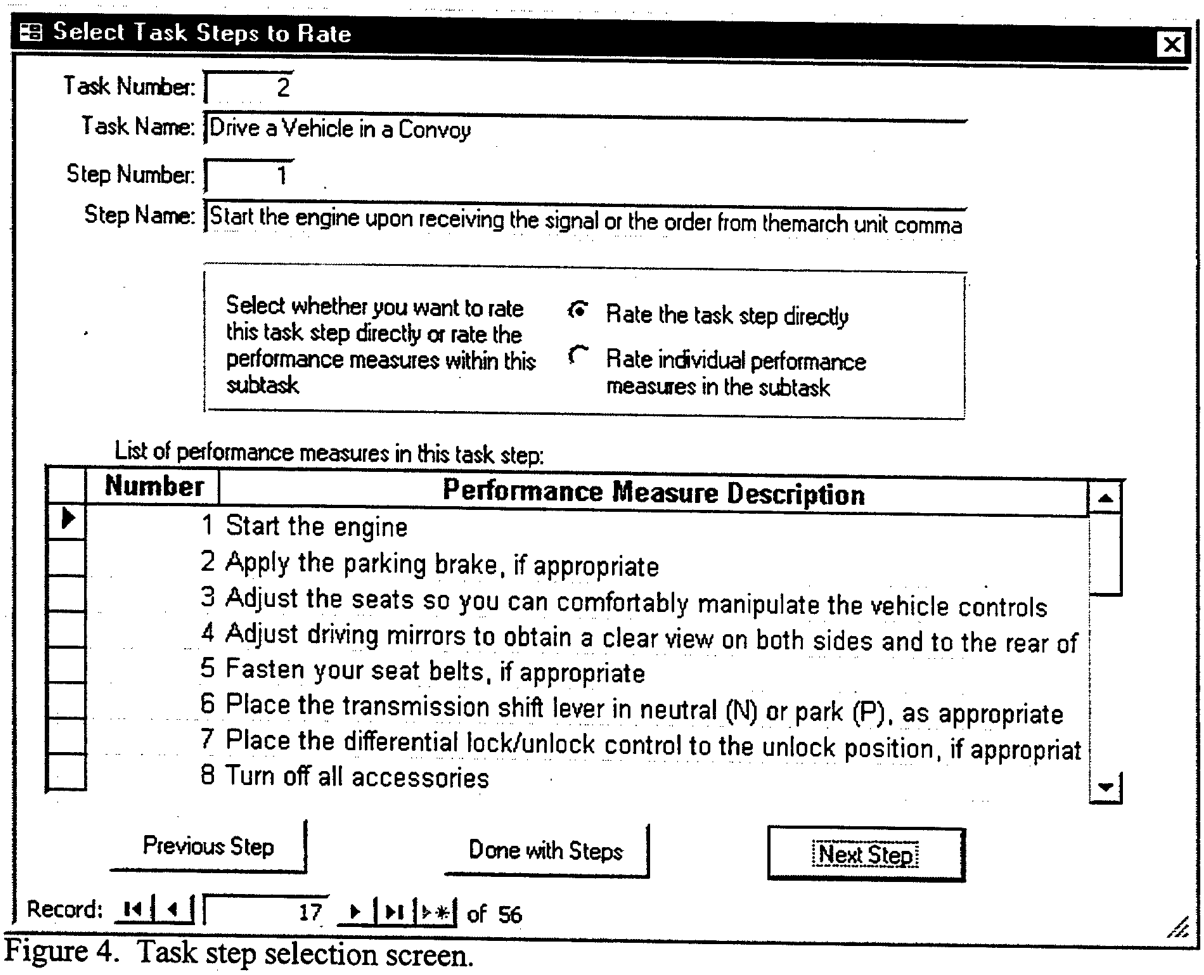
Figure 3. Task selection screen.

The task step selection form (Figure 4) looks essentially the same as the task selection form, and the procedure and rationale for selecting task steps, rather than performance measures, is also the same. The form presents the task steps for all tasks that were not selected to be rated at the task level. The user is asked the following question.

> *Select whether you want to rate this task step directly or rate the performance measures within this task step.*
>
> - *Rate the task step directly.*
> - *Rate the individual performance measures in the task step.*

As is the case at the task level, the user may decide to rate this step directly if the performance measures are similar in their technology requirements, if the user understands the step as a unit very well, or if it is obvious to the user that the step is easily supported by virtual environment technology. Otherwise, the user should not rate the task step, and should make ratings at the performance measure level. For example, the performance measures shown in Figure 4 are all procedures conducted in the cab of the truck. Since these are all similar, they can be rated together, as indicated in the figure.

After the user has made a selection on the first task step, he or she should proceed to the next step until all have been completed. When the task step selections are complete, the user should press the button labeled "Done with Steps" to continue with the analysis. Doing so creates a working data table that will be used for rating the selected tasks, task steps, and performance measures.



Task Number: 2
Task Name: Drive a Vehicle in a Convoy
Step Number: 1
Step Name: Start the engine upon receiving the signal or the order from themarch unit comma

Select whether you want to rate this task step directly or rate the performance measures within this subtask

● Rate the task step directly
○ Rate individual performance measures in the subtask

List of performance measures in this task step:

Number	Performance Measure Description
1	Start the engine
2	Apply the parking brake, if appropriate
3	Adjust the seats so you can comfortably manipulate the vehicle controls
4	Adjust driving mirrors to obtain a clear view on both sides and to the rear of
5	Fasten your seat belts, if appropriate
6	Place the transmission shift lever in neutral (N) or park (P), as appropriate
7	Place the differential lock/unlock control to the unlock position, if appropriat
8	Turn off all accessories

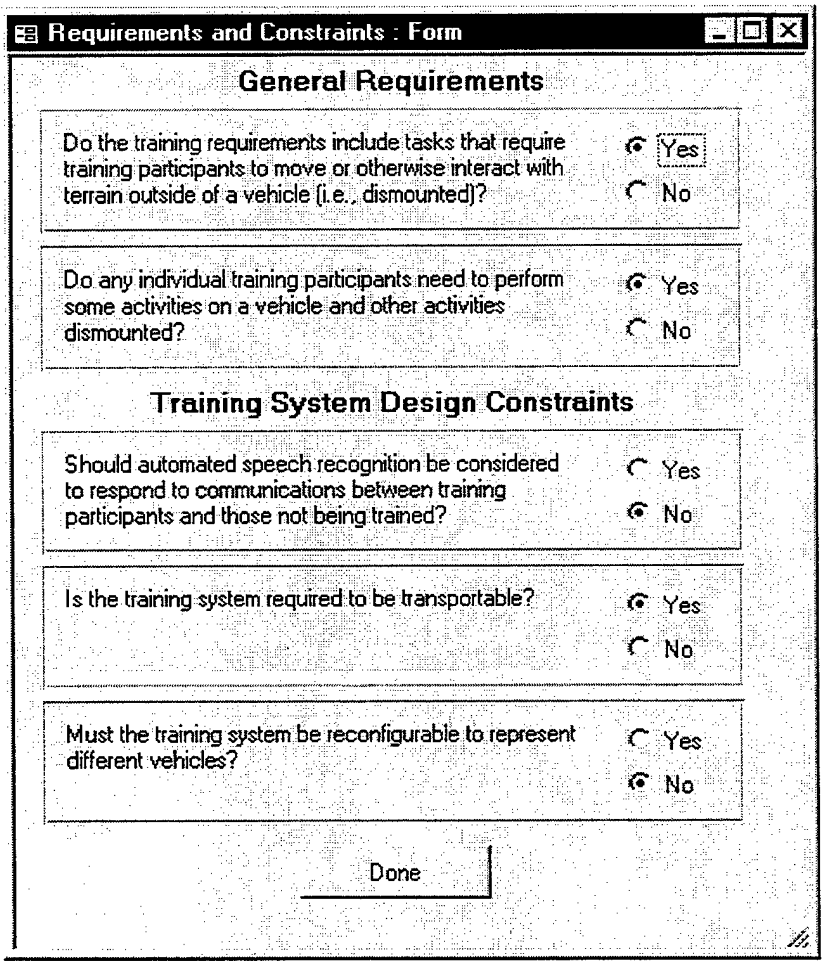
Record: 17 of 56

Figure 4. Task step selection screen.

It should be noted that the ratings made in the next step are not saved from the working data table to the permanent data tables until the results are calculated. Consequently, the user should not change the selection of tasks, task steps, and performance measures before the ratings are completed and results calculated. A user who wants to change the selection of tasks after making some ratings should select the "Calculate Results" button, which will save the ratings, then revise the task selection. This characteristic is obviously a limit of the demonstration that would be corrected in an operational version of this method.

Review Requirements and Design Constraints

Requirements and design constraints represent two steps of the STRIVE method. These steps require the user to answer the questions shown in Figure 5. The second question is disabled if the answer to the first question is "No." It is possible that the user will not know the answers to the questions assessing design constraints. In that case, it is probably best to make the most general assumptions, that is, assume that speech recognition is required, and that there are no requirements for a transportable or reconfigurable system. If these assumptions are made, then they can be changed later without requiring additional ratings. However, if the opposite assumption is made regarding speech recognition, then a change of that assumption would require the user to rate all task elements regarding that technology dimension.



General Requirements

Do the training requirements include tasks that require training participants to move or otherwise interact with terrain outside of a vehicle (i.e., dismounted)? ● Yes ○ No

Do any individual training participants need to perform some activities on a vehicle and other activities dismounted? ● Yes ○ No

Training System Design Constraints

Should automated speech recognition be considered to respond to communications between training participants and those not being trained? ○ Yes ● No

Is the training system required to be transportable? ● Yes ○ No

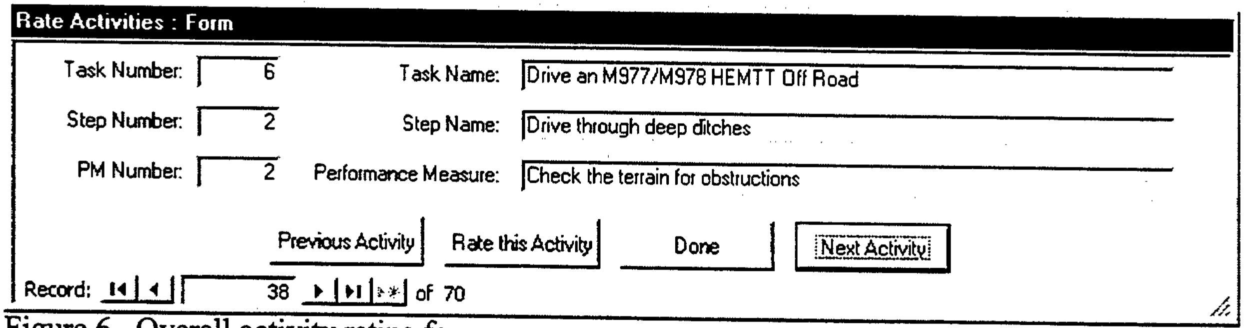
Must the training system be reconfigurable to represent different vehicles? ○ Yes ● No

Figure 5. General requirements and constraints questions.

Rate Selected Activities

When this option is selected from the main menu, the user sees the overall rating form shown in Figure 6. This form shows the name of the task, task step, or performance measure being rated. For example, Figure 6 shows that the activity being rated is the second performance measure in the second step of the sixth task. If a rating were being made at a more aggregated level, the performance measure and/or task step name would be blank.

Selecting the option to "Rate this Activity" brings up a series of questions about the activity. The specific questions were described in the previous section and will not be shown here. Questions only cover the technology dimensions that are consistent with the requirements and design constraints. After answering each question, the user selects the button labeled "Done" and continues to the next question. When all questions for a particular activity have



Task Number: 6
Task Name: Drive an M977/M978 HEMTT Off Road
Step Number: 2
Step Name: Drive through deep ditches
PM Number: 2
Performance Measure: Check the terrain for obstructions
Record: 38 of 70

Figure 6. Overall activity rating form.

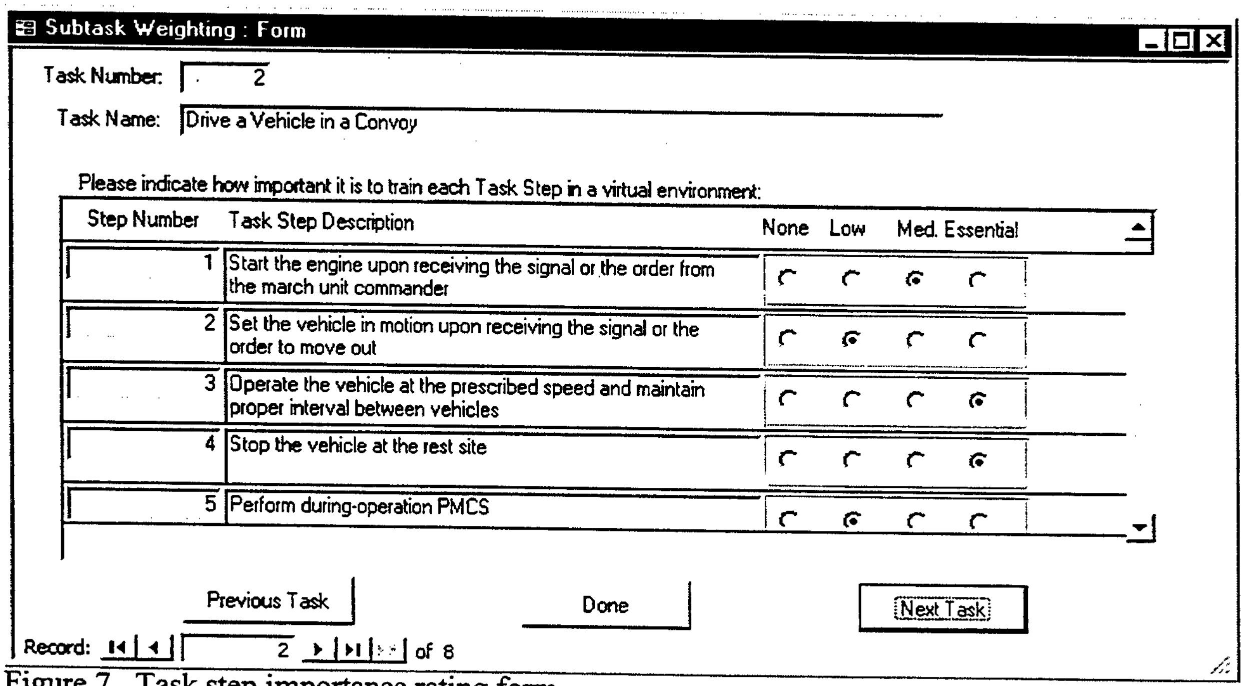
been selected, the user continues to the next activity, until all have been rated. It is important to answer all questions, because the method gives a score of "0" to all activities with missing data.

Rate Task Step Importance

The importance of training a task step in a virtual environment is assessed using the form shown in Figure 7. A task step may receive a low rating if it is not critical to the task or if it can be trained using some other method. As the figure shows, the user selects one of the radio buttons for each step. When all of the steps in a task have been rated, the user goes to the next task, until the user has rated the importance of all task steps. The user then selects the button labeled "Done" to continue the analysis. At this point in the analysis, all the necessary input data have been collected.

Calculate Results

Selecting this option initiates a procedure that calculates all support scores, copies scores to more detailed levels, aggregates scores to less detailed levels, and copies both ratings and



Task Number: 2
Task Name: Drive a Vehicle in a Convoy

Please indicate how important it is to train each Task Step in a virtual environment:

Step Number	Task Step Description	None	Low	Med.	Essential
1	Start the engine upon receiving the signal or the order from the march unit commander	○	○	●	○
2	Set the vehicle in motion upon receiving the signal or the order to move out	○	●	○	○
3	Operate the vehicle at the prescribed speed and maintain proper interval between vehicles	○	○	○	●
4	Stop the vehicle at the rest site	○	○	○	●
5	Perform during-operation PMCS	○	●	○	○

Record: 2 of 8

Figure 7. Task step importance rating form.

scores from the working data table to the appropriate permanent data table. The calculations take between a few seconds and a minute to run, depending on the speed of the machine. There are no displays associated with this option.

It should be noted that technology dimensions that are not being considered are given a score of "3" (high support). This assignment ensures that the excluded dimension will not be considered in the overall score, which is based on the minimum technology dimension score.

View Results

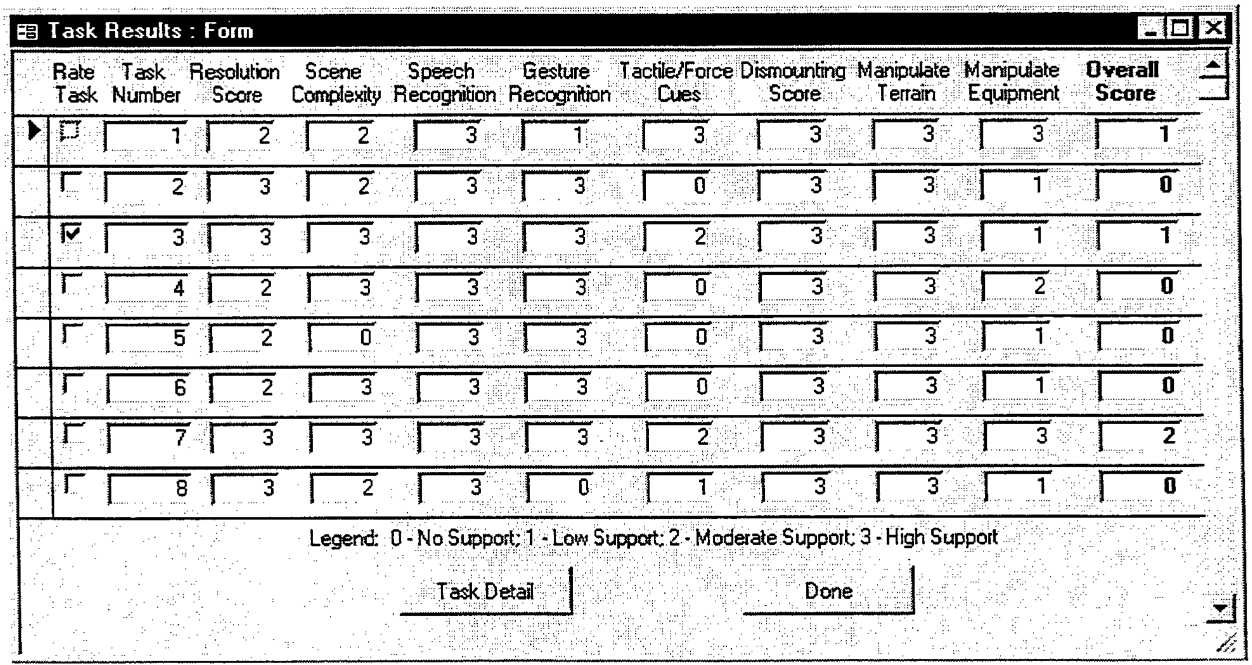
Selecting this option brings up a display of the results of the analysis. Figure 8 shows the overall results for all tasks, while Figure 9 shows the results for the task steps in Task 1. These figures illustrate some of the displays that could be produced. Additional displays and reports are also possible, but were not developed as a part of the demonstration.

Quit

The final option closes the main menu form, thus completing the analysis. The STRIVE database can then be closed and Microsoft Access exited.

Example Problems

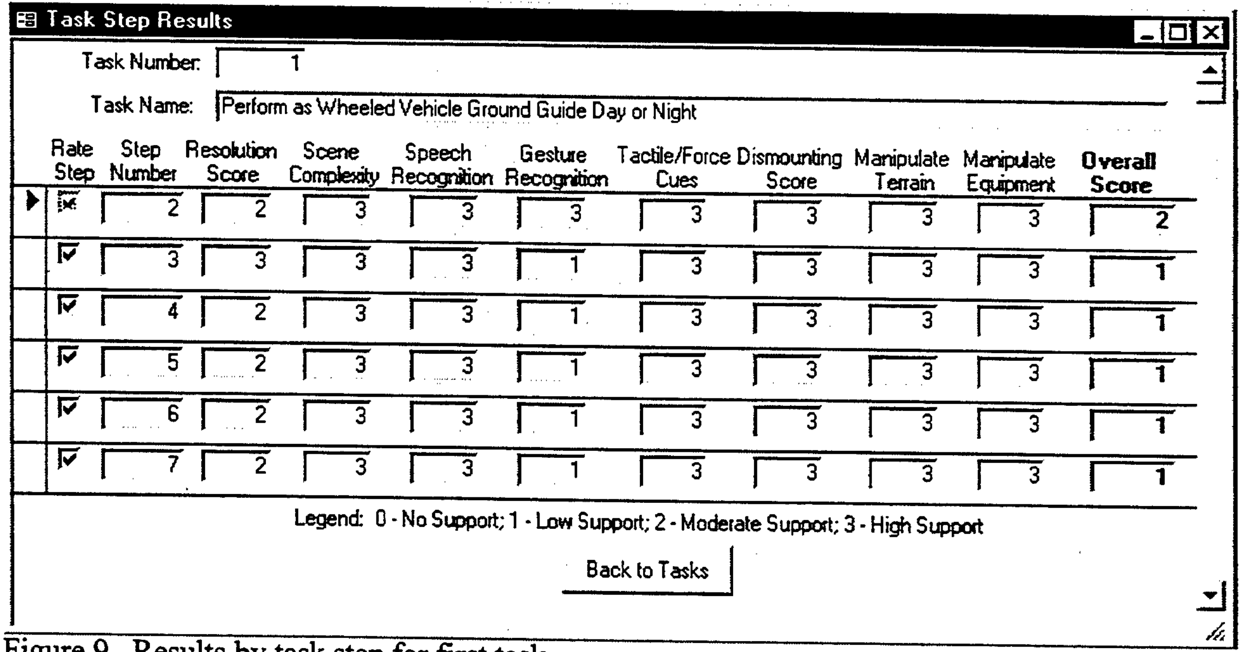
The two examples chosen to illustrate the method differ in several respects. The AVCATT-A is based on an existing system requirement to simulate collective aviation tasks. There is no similar requirement for the HEMTT, which is a family of large trucks. Furthermore, a training system for this vehicle would focus on individual tasks. Nevertheless, the same basic procedure was used to select tasks to be incorporated into the examples, as described in the following discussion.



Rate Task	Task Number	Resolution Score	Scene Complexity	Speech Recognition	Gesture Recognition	Tactile/Force Cues	Dismounting Score	Manipulate Terrain	Manipulate Equipment	**Overall Score**
☐	1	2	2	3	1	3	3	3	3	**1**
☐	2	3	2	3	3	0	3	3	1	**0**
☑	3	3	3	3	3	2	3	3	1	**1**
☐	4	2	3	3	3	0	3	3	2	**0**
☐	5	2	0	3	3	0	3	3	1	**0**
☐	6	2	3	3	3	0	3	3	1	**0**
☐	7	3	3	3	3	2	3	3	3	**2**
☐	8	3	2	3	0	1	3	3	1	**0**

Legend: 0 - No Support; 1 - Low Support; 2 - Moderate Support; 3 - High Support

Figure 8. Task results display.



Task Number: 1

Task Name: Perform as Wheeled Vehicle Ground Guide Day or Night

Rate Step	Step Number	Resolution Score	Scene Complexity	Speech Recognition	Gesture Recognition	Tactile/Force Cues	Dismounting Score	Manipulate Terrain	Manipulate Equipment	**Overall Score**
☑	2	2	3	3	3	3	3	3	3	2
☑	3	3	3	3	1	3	3	3	3	1
☑	4	2	3	3	1	3	3	3	3	1
☑	5	2	3	3	1	3	3	3	3	1
☑	6	2	3	3	1	3	3	3	3	1
☑	7	2	3	3	1	3	3	3	3	1

Legend: 0 - No Support; 1 - Low Support; 2 - Moderate Support; 3 - High Support

Figure 9. Results by task step for first task.

Aviation Combined Arms Tactical Trainer – Aviation Reconfigurable Manned Simulator (AVCATT-A)

The AVCATT-A is to be a networked virtual environment simulator providing collective and combined arms training and rehearsal in a simulated battlefield environment. It will be used by both Active and Reserve Component aviation units worldwide. Because it will be interoperable with CCTT and other HLA-compliant systems, it will be possible to simulate combined arms operations with a variety of ground vehicles. In addition, AVCATT-A will have the capability to represent attack, reconnaissance, cargo, and utility aircraft, SAF workstations, AAR capability, a battlemaster control console, and workstations for ground maneuver, fire support, close air support, logistics, battle command, and engineer role players.

Requirements for the AVCATT-A are found in the following documents:

- *Aviation Combined Arms Tactical Trainer and the Aviation Reconfigurable Manned Simulator (AVCATT-A): Operational Requirements Document.* 12 April 1999 Revision.
- *System Requirements Document: Aviation Combined Arms Tactical Trainer – Aviation Reconfigurable Manned Simulator (AVCATT-A).* Orlando, FL: Simulation, Training, and Instrumentation Command (STRICOM), 22 October 1999.

AVCATT-A is required to be a mobile, transportable, trailerized system. This requirement places constraints on both the visual display system and the motion cueing system that are used. Because of the limited space allowed in a trailer, dome display systems and platform motion systems will not be feasible. Although this constraint might limit the potential capability of the system, the characteristics of the tasks will determine whether this limitation has any practical significance.

A key feature of AVCATT-A is its planned use of reconfigurable manned simulators. Each training device will be able to simulate the AH-1F Cobra, AH-64A Apache, AH-64D Longbow Apache, RAH-66 Comanche, OH-58D Improved/Improved Optimized/Digitized, UH-60A/L/X Blackhawk, UH-1H Iroquois, CH-47D Chinook, Ch-47D Improved Cargo Helicopter, and Light Utility Helicopter aircraft. This flexibility appears to require a helmet-mounted display system. The SRD recognizes this likelihood an it describes the requirement for a helmet-mounted display in considerably greater detail than the requirement for a direct-view display.

In addition to these constraints, both the ORD and the SRD give direct requirements for some of the cues that must be represented, responses sensed, and activities supported. For example, the ORD states that sounds must be represented in the appropriate quadrant of the crew station, and that cockpit indications must include vibration cues. The visual range is required to be sufficient for the pilot to make accurate estimates of distance, velocity, and height.

The following documents provide information on aviation operations and tasks:

- FM 1-112. *Attack Helicopter Operations.* 2 April 1997
- FM 1-113. *Utility and Cargo Helicopter Operations.* 25 June 1997.
- FM 1-114. *Air Cavalry Squadron and Troop Operations.* 1 February 2000.
- ARTEP 1-112-MTP. *Mission Training Plan for the Attack Helicopter Battalion.* 30 March 2000
- ARTEP 1-113-MTP. *Mission Training Plan for the Utility Helicopter Battalion.* 30 March 2000
- ARTEP 1-114-MTP. *Mission Training Plan for the Air Cavalry/Reconnaissance Squadron and Troop.* 30 March 2000.

The tasks were selected from the three ARTEP MTPs. There is considerable overlap between these three documents. MTP112 and MTP114 are identical, containing the same 119 tasks with the same titles, ID numbers, task steps, performance measures, and supporting individual tasks. MTP113 contains 107 tasks. Of these, 99 tasks are identical with both MTP112 and MTP114, while 8 tasks are specific to the Utility Battalion. In addition, 12 tasks that appear in both the Attack Battalion and Air Cavalry Squadron are not included in the Utility Battalion. There are two elements only found in the Utility Battalion (CEWI Platoon and Pathfinder Platoon). Each of these elements is associated with one task.

Tasks were selected for the demonstration to illustrate a variety of task characteristics. Some of the selected tasks present a minimal challenge to virtual environment technology, while others present a more substantial challenge. Some appeared to be relatively homogeneous, so that they might be rated at the task or task step level, while others were more heterogeneous and might need to be rated at the performance measure level. Although six of the selected tasks were taken from the AVCATT-A requirements, three were not on that list. These three presented a significant challenge to the capabilities of virtual environment technology. The following nine selected tasks contain 53 task steps and 197 performance measures.

- Conduct downed aircrew recovery operations (01-2-0108.0NRC)
- Conduct deliberate attack (01-2-0211.01-0NRC)
- Participate in the staff planning process (S3) (01-101301.01-0NRC)
- Conduct air Volcano operations (01-2-1334.01-0NRC)
- Conduct aviation urban operations (01-1-1343.01-0NRC)
- Provide pathfinder support (01-3-1353.01-0NRC)
- Conduct battle handover/relief on station (01-2-2044.01-0NRC)
- Conduct air movement operations (01-2-5103.01-0NRC)
- Conduct air assault operations (01-2-5105.01-0NRC)

Tasks were rated by an Army civilian working in the Directorate of Training, Doctrine, and Simulation. The rater had extensive experience with designing and programming simulation-based aviation training, and was thoroughly familiar with the tasks and the AVCATT-A requirements. Although the rater was not able to rate all tasks due to time constraints, he was able to give feedback on the overall procedure, the organization of task elements, and the questions that were asked to address technology requirements.

Heavy Expanded Mobility Tactical Trucks (HEMTT)

The HEMTT is a large truck that provides transport capabilities for resupply of combat vehicles and weapon systems. There are five versions of this vehicle:

- The M977 is a cargo truck used for resupply of ammunition between the Field Artillery Ammunition Support Vehicle and the Ammunition Supply Points. It is also used to resupply the Armored Forward Area Rearm Vehicle

- The M978 is a 2,500-gallon tanker used to move fuel forward from battalion trains to preselected areas close to the Forward Line of Troops where combat vehicles will withdraw to refuel.

- The M984 is a wrecker-recovery vehicle used to tow a wide variety of loads and perform vehicle recovery.

- The M983 is a tractor used to transport Pershing II missiles and Patriot missile system semitrailers.

- The M985 is a cargo truck with material handling crane used for resupply of the Multiple Launch Rocket System (MLRS).

This example does not come from an existing device requirement; consequently, there is no ORD or SRD. The following documents contain the potential training requirements and other activities conducted by the HEMTT operator.

- TM 9-2320-279-10-1. *Operator's Manual Volume No. 2 M77 Series 8 X 8 Heavy Expanded Mobility Tactical Trucks (HEMTT)*. 15 June 1987.
- STP 55-88M12-SM. *Soldier's Manual MOS 88M Motor Transport Operator Skill Levels 1 and 2*. 23 December 1993.
- TC 21-305-1. *Training Program for the Heavy Expanded Mobility Tactical Truck (HEMTT)*. 3 October 1995.

We used two documents – the Soldier's Manual and the TC 21-305-1 – as sources for tasks. These documents have very different uses and formats. The Soldier's Manual is very general and often somewhat superficial. The TC is primarily intended as an Instructor's Guide. It is specific to the HEMTT and is much more detailed, but it is instruction oriented rather than field-performance oriented. The HEMTT domain of tasks is much smaller than the domain for the AVCATT. We identified fewer than 30 tasks in the two sources. As was the case for the AVCATT-A, there is considerable overlap between the two sources of tasks, and also within each individual source documents, in that many tasks subsume other tasks.

The task terminology is not consistent between the two sources, nor is it the same as the terminology used in the Aviation MTPs. The SM uses the word "task" as the highest order of designate but does not identify task elements by name. In general, the SM goes down two levels below task in the Training Information Outline. The words "Performance Measures" are reserved for the Evaluation Guide, where they appear to be used for what are called "steps" in the MTPs. Likewise, the word "task" is used in the TC, but the lower levels in the hierarchy are not identified by name.

Tasks were selected for the HEMTT example with the same considerations used for the selections for the AVCATT-A example. The following eight selected tasks contain 66 steps and 345 performance measures (although these terms are not used in the source documentation).

- Perform as wheeled vehicle ground guide day or night (551-721-1384)
- Drive a vehicle in a convoy (551-721-1359)
- Identify major components, cab controls, instruments, and indicators (HEMTT) (derivative of 551-721-1352)
- Operate engine brake (jake brake) (derivative of 551-721-1366)
- Drive the HEMTT on the road (primary or secondary) (derivative of 551-721-1366)
- Drive an M977/M978 HEMTT off road (derivative of 551-721-1360)
- Drive the HEMTT at night (derivative of 551-721-1366)
- Operate an M977 HEMTT crane (derivative of 551-721-1407 and 551-721-1352)

The eight HEMTT tasks were rated by one of the authors, who has moderate familiarity with their content. Task documentation was used extensively to make the ratings, which took approximately 12 hours. The ratings and resulting scores are incorporated in the STRIVE demonstration.

Rater Feedback Regarding Example Problems

The raters for the two example problems provided feedback regarding several aspects of the STRIVE methodology. Some of the comments concerned issues that are not unique to STRIVE. For example, both expressed the opinion that the task descriptions could be improved. The AVCATT-A rater suggested that the collective tasks that were rated, which were taken from the ARTEP MTP, should be linked to related individual tasks. The rater anticipated that use of individual tasks would produce more accurate ratings because many of the requirements being rated differ among the individuals in a unit who are being trained. This criticism would also apply to TPS code analysis or other methods that are based on ratings of ARTEP MTP tasks.

The ratings of the HEMTT were already applied to individual tasks. However, the rater commented that the level of detail in the description of the tasks was inconsistent and often insufficient to support a rating. Because the rater was not expert in the operation of the HEMTT, he had to rely on the documentation to provide the information required for the ratings. Although the detail of documentation is likely to be a general problem, it would probably have less impact for a more experienced rater, who could rely on experience to compensate for deficiencies in task documentation.

Other problems noted by the raters were specific to the STRIVE methodology and demonstration. Both raters had some difficulty in determining the appropriate level at which to make ratings. On the basis of this feedback, we anticipate that the user interface used to obtain these judgments should be changed and that the instructions for operation of the method should be enhanced. This report includes a more detailed description of the process used to select the task elements that will be rated than was available to the raters. A more effective user interface might incorporate task and task-step selection in a single process, rather than use the two-step procedure that was incorporated in the demonstration.

Finally, the AVCATT-A rater made two suggestions regarding the questions used to assess the requirements for visual display resolution. First, an additional activity that would require high resolution is identifying targets. Target identification requires more detail than target detection. Consequently, identification of a target may require greater resolution than detection, even if the target to be identified is closer. Second, the assessment procedure should consider the possibility that visual activities will use a magnified sight. The level of magnification should be considered in assessing the resolution requirement. The AVCATT-A rater made a final suggestion regarding questions addressing manipulation of terrain or equipment. These questions were asked for each task element that was rated, when they could have been answered once for the tasks as a whole.

The feedback from the raters provides guidance for future implementation and enhancement to the STRIVE methodology. Each of these comments can be addressed with specific changes to the procedures without changing the overall methodology.

IMPLEMENTATION OF METHOD IN TRAINING DEVICE DEVELOPMENT PROCESS

To be useful, the STRIVE methodology must be integrated into the training device development process, which is governed by TRADOC Regulation 350-70 (1999). In addition, implementation of the method will require the development of capabilities that are not included in the demonstration, namely procedures to manage task requirement and technology capability information.

Incorporating STRIVE in TADSS Development

The goal of the STRIVE methodology is to aid in establishing requirements for designing virtual environment training systems. These requirements are established early in the TADSS design process, as illustrated in Figure 10. Of particular interest in this process is the ORD, which is used to initiate the development and procurement of a TADSS. Our review of the documentation for the development of AVCATT-A indicated that the analyses that occur after the development of the ORD are at a greater level of detail than could be supported by the STRIVE methodology. For example, the SRD for AVCATT-A (1999) was supported by fidelity analyses that examined each individual cockpit control and display for all aircraft required to be represented by the AVCATT-A system. This level of detail is substantially beyond the level that was envisioned for the STRIVE methodology.

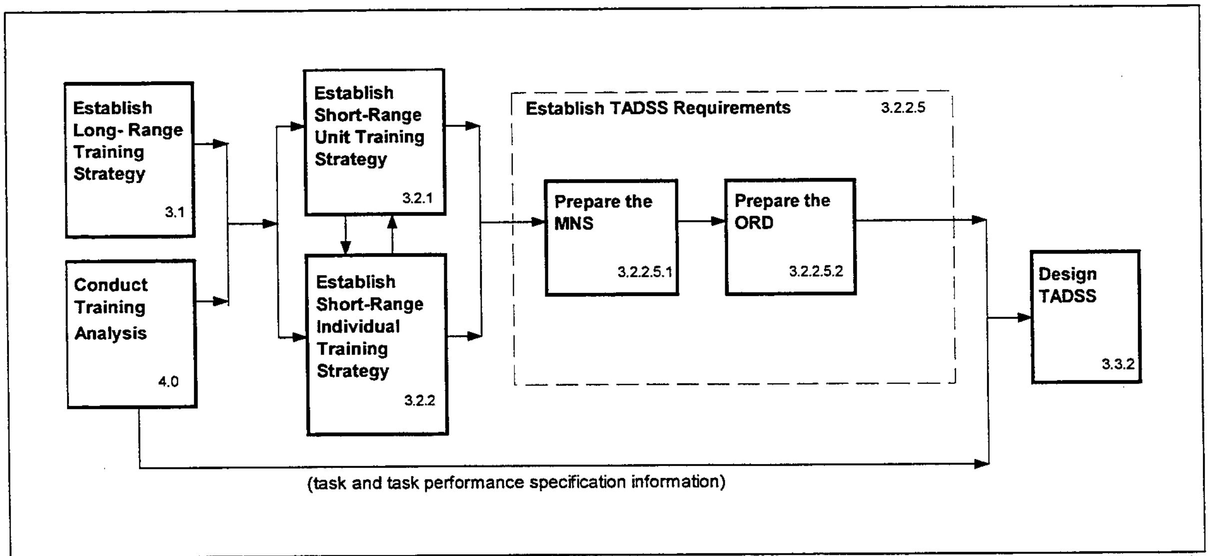


Figure 10. Process for establishing TADSS Requirements (from TRADOC Regulation 350-70).

On the other hand, because the ORD specified the tasks that are required to be represented by the TADSS, the development of this document could be aided by the application of an analysis method such as STRIVE. The design and development activities that occur before the approval of the ORD are termed concept exploration and definition. As Figure 10 illustrates, the ORD is supported by training analysis, the development of training strategy, and a Misison Needs Statement (MNS). Because the MNS is required only when the development is in response to new missions, we will not discuss it here. The training strategy is a general description of the training methods to be used and the resources required to implement these methods.

The STRIVE method can support the development of the ORD for a virtual environment training system in several ways. First, it can aid in the selection of the individual or collective tasks that are included in the operational requirements. The application of STRIVE can help ensure that the tasks assigned to virtual environment training are realistic given the current technological capabilities. Furthermore, STRIVE can help in the development of a coherent training strategy that coordinates training in live, virtual, and constructive environments. STRIVE will be one of several tools required to develop such a strategy. Other tools will be needed to evaluate other training environments and to address and cost and training efficiency considerations.

The best option for the implementation of STRIVE is as a component in a suite of analysis methods to support the concept exploration and definition process. This system would be analogous to the Automated System Approach to Training (ASAT) and the Standard Army Training System (SATS), each of which combines several tools for training design and management. Currently, few of the tools that would be integrated with STRIVE are available. One possibility is the Training Mix Model (Djang, Butler, Laferriere, and Hughes, 1993), which can provide guidance for allocating tasks to training environments. Other tools might involve rough order-of-magnitude cost estimation, early estimation of training effectiveness (as opposed to task support), and determination of the most appropriate level of technical sophistication. All of these tools would need to require data at a level of detail that is consistent with the early phase in the development process.

Alternatively, STRIVE could be implemented as a component of an existing system. The best choice for such as system seems to be ASAT. This option would have some benefits in facilitating the management of task data, as described in the following section. However, the focus of STRIVE on the TADSS development process would make it substantially different from the other tools that make up ASAT. A final possibility is the independent implementation of STRIVE. While, this option has benefits in the short term, the greatest value of the methodology will be obtained when it is combined with other compatible tools.

Requirements for Task Requirement and Technology Capability Data

Procedures for managing task and technology data were not included in the STRIVE demonstration. Each of these data management capabilities would be required for the implementation of an operational version of STRIVE.

Task Data

We anticipate that the operational version of STRIVE would obtain task data from the Reimer Digital Library Data Repository (RDL DR). The RDL DR contains a relational task database that currently includes information about over 26,000 collective and individual tasks. Task data included in the RDL DR were developed by the ASAT system. Currently, the relational information can be accessed only by other systems, including ASAT and SATS. Direct queries regarding specific individual or collective tasks are answered in hypertext markup language (HTML) format. Use of this source of task data would allow STRIVE to use the most current task definitions and would eliminate most of the clerical effort required to organize and enter task information. It also has the potential to provide links between collective and individual tasks and thus could satisfy one of the criticisms of the raters. Because the task data are produced by the ASAT system, they would follow the rules of consistency established by that system, which should reduce or eliminate problems of inconsistent decomposition of tasks.

An operational version of STRIVE would require direct access to the relational task data in the RDL DR, to obtain task step and performance measure information. One potential way to accomplish this link would be to incorporate STRIVE within ASAT. This alternative would eliminate the need to develop separate task data management capabilities, because ASAT already includes the capability to import and export task information. Furthermore, incorporating STRIVE in ASAT would allow proponents to use the method to identify the role of virtual environment technology in the training strategy, as well as the operational requirements necessary to eliminate training deficiencies. In this way STRIVE would support the proponents in developing their input to the ORD.

However, the capabilities of ASAT are not oriented to the development of new training devices. ASAT is focused on creation and management of task information and on the development of specific products, such as Mission Training Plans, Drill Books, Soldier Training Publications, Training Support Packages, and Lesson Plans. STRIVE, on the other hand, is specifically oriented toward the TADSS-development process. Consequently, STRIVE would represent a new category of functionality for ASAT, and may best be developed as an independent capability that would be integrated with other tools used in the TADSS-development process.

Technology Capability Data

In many respects, management of technological capability information is more difficult than the management of task data. Advances in the capabilities of relevant virtual environment technologies occur constantly and are results of the research and development efforts of many independent corporations and other organizations. The speed of technological advancement implies that capabilities must be monitored closely to ensure that they are accurate. The existence of many independent developers implies that it will be necessary to survey a large number of sources to accurately characterize capabilities. Furthermore, while tasks are relatively independent entities, technology offerings often compromise performance on several dimensions to provide a useful capability at a reasonable price. For example, the visual field of view of a particular image generation system may be reduced to allow higher resolution at the center of the display. Similarly, the complexity of a moving model in an image generation system may be reduced so that more models can be displayed simultaneously.

Development of a detailed procedure to assess technology capability will take some effort. The resulting procedure should have the following components:

- Standardization of the reporting of technology capabilities;

- A periodic survey of technology vendors to assess current and planned future technology capabilities;
- Publication of the results of the survey for comment from vendors and other members of the training and simulation community.

Use of a procedure with these three features will ensure that technology information that is compared to task requirements is both consistent and accurate.

SUMMARY AND CONCLUSIONS

Virtual environment technology has been successfully applied to individual, crew, and collective training. The history of applications of the technology has provided some information regarding the kinds of activities that can be performed in a virtual environment and the kinds of activities that can't. The STRIVE method attempts to summarize the results of earlier evaluations and other analyses in a form that can be used to guide the design of new training systems.

The method extends the existing TPS code analysis so that it can be applied before a training system has been designed. This extension requires a type of behavioral analysis of tasks to be trained instead of the direct rating of task performance support that is a part of TPS code analysis. Because the incorporation of a behavioral analysis can increase rater workload, several features were incorporated into the design that make the analysis as efficient as possible.

- The technology dimensions are limited to those that may present a challenge to the capabilities of virtual environment technology.
- The technology dimensions are further limited to those that present the greatest problem in the training domain of interest (e.g., are dismounted soldiers involved) and to those that are consistent with any training system design constraints (e.g., should automated speech recognition be used).
- Ratings are made at the task and task step level, rather than the performance measure level, whenever possible. The heterogeneity of the elements of a task does not permit meaningful ratings for all tasks. However, workload can be decreased by increasing the level of aggregation of the activities to be rated, when this is feasible.

The resulting procedure assesses the capability of virtual environment technology to support task performance based on SME judgments of selected cues and responses needed to perform task activities. The user of STRIVE selects the level of detail at which ratings will be made for each task, describes selected requirements of the training domain and design constraints, rates the cues and responses of the selected task elements, and assesses task step importance. Based on these input data, the method calculates a score representing the extent to which the task elements can be supported by virtual environment technology. The scores are migrated to lower levels and aggregated to higher levels up to the task level.

The feasibility of the procedure was demonstrated with two example problems from considerably different domains. One of these, the AVCATT-A, represented a collective, combined arms training system for which there is an existing system requirement. The other example, the HEMTT, represents an individual training domain with no training device requirement currently expressed. To facilitate the development of the examples, an automated demonstration was developed using Microsoft Access97. Although the demonstration is not operational software, it represents all method functions and implements all selection procedures and calculations.

Several issues regarding the capability of virtual environment technology could not be solved by this effort. These issues present challenges to future research in this area.

The relative paucity of training evaluation studies for virtual training systems limits both the level of detail and the accuracy of the information in the STRIVE model. The difficulty and expense of assessing training effectiveness, particularly for collective training systems, has limited the information that is available to guide the design of future training systems and to establish the needs for future technology development. Although it seems clear that an investment in quality evaluation data will yield returns in future development efficiency, this knowledge has not been sufficient to encourage the careful evaluation of emerging training systems.

Lacking data on training effectiveness, STRIVE follows the approach of TPS code analysis to focus on task performance support. However, often, the questions that must be decided are not ones of possibility, but represent concerns about affordability, and cost-effectiveness. The substantial capabilities of virtual environment technology make many types of simulations possible in both the individual and collective training arenas. However, it is not clear that just because a training requirement can be met in a virtual environment, that it should be met in that environment. To answer this question properly requires consideration of both the cost required to meet the requirement as well as the training effectiveness.

The STRIVE method has focused on existing technological capabilities. Forecasting future capabilities of technology has been difficult in some areas. While it is possible to develop a reasonable projection of the cost of memory and processor speed that would be available to produce an image generation system with certain capabilities, projections in other areas are more problematic. This issue is made more complex by the fact that requirements for military training systems may be a major force for the development of some technologies, while in other areas, the military may need to capitalize on developments in civilian technology applications. Forecasting future technology capabilities is a difficult problem that will require considerable effort to solve.

The capabilities of the STRIVE demonstration illustrate the potential for this method. However, realizing this potential will require the development of an operational implementation of the procedure. This implementation should incorporate other features, such as management of task and technology capability data, as well as the incorporation of a more robust user interface.

REFERENCES

Boldovici, J. (1993). *Simulator Motion* (ARI Technical Report No. 961). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Burkett, M.L. & Mullen, W.J. (2000). *Live Fire Futures* (Draft ARI Study Report 99-02). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Burnside, B.L. (1990). *Assessing the capabilities of training simulations: A method and simulation networking (SIMNET) application* (Research Report 1565). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A226-354)

Close Combat Tactical Trainer (CCTT) Accreditation Report (1999). Author unknown.

Department of the Army (1 February 2000). *Air Cavalry Squadron and Troop Operations* (FM 1-114). Washington, DC: Author.

Department of the Army (2 April 1997). *Attack Helicopter Operations* (FM 1-112). Washington, DC: Author.

Department of the Army (30 September 1990). *Battle focused training* (FM 25-101). Washington, DC: Author.

Department of the Army (30 March 2000). *Mission Training Plan for the Air Cavalry/ Reconnaissance Squadron and Troop* (ARTEP 1-114-MTP). Washington, DC: Author

Department of the Army (30 March 2000). *Mission Training Plan for the Attack Helicopter Battalion* (ARTEP 1-112-MTP). Washington, DC: Author.

Department of the Army (30 March 2000). *Mission Training Plan for the Utility Helicopter Battalion* (ARTEP 1-113-MTP). Washington, DC: Author.

Department of the Army (15 June 1987). *Operator's Manual Volume No. 2 M77 Series 8 X 8 Heavy Expanded Mobility Tactical Trucks (HEMTT)* (TM 9-2320-279-10-1). Washington, DC: Author

Department of the Army (23 December 1993). *Soldier's Manual MOS 88M Motor Transport Operator Skill Levels 1 and 2* (STP 55-88M12-SM). Washington, DC: Author.

Department of the Army (3 October 1995). *Training Program for the Heavy Expanded Mobility Tactical Truck (HEMTT)* (Training Circular 21-305-1). Washington, DC: Author.

Department of the Army (25 June 1997). *Utility and Cargo Helicopter Operations*. (FM 1-113). Washington, DC: Author.

Djang, P.A., Butler, W.A., Laferriere, R.R., & Hughes, C.R. (1993). *Training mix model.* TRAC-WSMR-TEA-93-035. White Sands Missile Range, NM: TRADOC Analysis Center. (ADB178428)

Ford, L.A., & Campbell, R.C. (1997). *Staff performance analysis: A method for identifying brigade staff tasks* ((RP-WATSD-97-14). Alexandria, VA: Human Resources Research Organization.

Gorman, P.F. (1991). The future of tactical engagement simulation. In D. Pace (Ed.), *Proceedings of the 1991 summer computer simulation conference* (pp. 1-9). Baltimore, MD: Society for Computer Simulation.

Jacobs, R. S., Crooks, W. H., Crooks, J. R., Colbunn, E., Fraser, 11, R. E., Gorman, P. F., Madden, J. L., Furness, III, T. A., & Tice, S.. E. (October, 1994*). Behavioral Requirements for Training and Rehearsal in Virtual Environments* (TR-1011), U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA. (ADA286311)

Keesling, J.W., King, J., & Mullen III, W.J. (1998). *Simulation training strategies for Force XXI final report.* Fort Knox, KY: BDM Federal, Inc.

Klein, G.A. (1982). *Use of comparison cases to predict training effectiveness of maintenance simulators.* Yellow springs, OH: Klein Associates.

Klein, G.S., Johns, P., Perez, R., & Mirabella, A. (1985). *Comparison-Based Prediction of Cost and Effectiveness of Training Devices: A Guidebook.* Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kribs, H.D., Simson, A.C., & Mark, L.J. (1983). *Automated instructional media selection (AIMS)* (NAVTRAEQUIPCEN 79-C-0104-1). San Diego, CA: Instructional Science and Development, Inc.

Meliza, L.L. (1993). *Simulation networking/training requirements relational database: User's guide* (Research Product 94-01). Orlando, FL: U.S. Army Research Institute for the Behavioral and Social Sciences, STRICOM, Orlando Field Unit, Florida Training Systems Research Division.

Muckler, F.A. & Finley, D.L. (1994). *Applying training system evaluation models to Army training: Volume I. Analysis of the literature* (ARL-TR-463). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Pallett, D.S., Fiscus, J.G., Garofolo, J.S., Martin, A., Pszsybocki, M.A. (1999). 1998 Broadcast News Benchmark Test Results. In *Proceedings of the DARPA Broadcast News Workshop.* Arlington, VA: Defense Advanced Research Projects Agency.

Rose, A.M., Martin A.W., & Yates, L.G. (1985). *Forecasting Device Effectiveness: Volume III. Analytic Assessment of Device Effectiveness Forecasting Techniques* (Technical Report 681). Alexandria, VA: U.S. Army Research for the Behavioral and Social Sciences.

Rose, A.M., Wheaton, G.R., & Yates, L.G. (1985a). *Forecasting Device Effectiveness: Volume I. Issues* (Technical Report 680). Alexandria, VA: U.S. Army Research for the Behavioral and Social Sciences.

Rose, A.M., Wheaton, G.R., & Yates, L.G. (1985b). *Forecasting Device Effectiveness: Volume II. Procedures* (Research Product 85-25). Alexandria, VA: U.S. Army Research for the Behavioral and Social Sciences.

SHERIKON, Inc. (1995). *Task performance support (TPS) codes: A means to allocate MTP tasks to simulation-based collective training events.* Orlando, FL: Author.

Simpson, H. (1995). *Cost-effectiveness analysis of training in the Department of Def*ense (TR 95-004). Seaside, CA: Defense Manpower Data Center.

Simulation, Training and Instrumentation Command (STRICOM) (22 October 1999). *System requirements document: Aviation Combined Arms Tactical Trainer – Aviation Reconfigurable Manned Simulator (AVCATT-A).* Orlando, FL: Author.

Simulation, Training and Instrumentation Command (STRICOM) (12 April 1999). *Aviation Combined Arms Tactical Trainer and the Aviation Reconfigurable Manned Simulator (AVCATT-A) Operational Requirements Document.* Orlando, FL: Author.

Simulation, Training and Instrumentation Command (STRICOM) (14 January 2000). *One Semi-Automated Forces (OneSAF) Operational Requirements Document (ORD) Version 1.1.* Orlando, FL: Author.

Sticha, P. J., Blacksten, H. R., Buede, D. M., Singer, M. J., Gilligan, E. L., Mumaw, R. J., & Morrison, J. E. (1990). *Optimization of simulation-based training systems: Model description, implementation, and evaluation* (Technical Report 896). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Sticha, P.J., Campbell, R.C., & Schwalm, S.R. (1996). *Virtual environment interface requirements for combat leader training and mission rehearsal* (FR-WATSD-96-15). Alexandria, VA: Human Resources Research Organization.

Strachan, I.W. (1998). *Jane's simulation and training systems, Eleventh Edition, 1998-99.* Surrey, UK: Jane's Information Group Limited.

Strachan, I.W. (1999). *Jane's simulation and training systems, Twelfth Edition, 1999-2000.* Surrey, UK: Jane's Information Group Limited.

U.S. Army Training and Doctrine Command (1999). Functional Concepts of the Standard Army Training System (SATS). In *Systems Approach to Training Management, Processes, and Products, TRADOC Regulation 350-70.* Fort Monroe, VA: Training and Doctrine Command.

APPENDIX A
SURVEY OF SELECTED ARMOR AND AVIATION TRAINING DEVICES

Our review of training systems at Ft. Knox and Ft. Rucker attempted to get information about how the systems were used as well as their technical capability. In general, detailed technical information about the systems was not available, but we were able to obtain a general idea about the capabilities from a demonstration or walk-though of each system. We were able to obtain written documentation describing how some of the systems were used. This information was supplemented by the discussions with trainers and managers at the sites.

Our discussion of each system will briefly outline the technical capability of major components, including visual, motion, weapons, and instructor support components. We will also summarize information we obtained from discussions and supporting documentation regarding the use of the systems and any perceived needs for additional capability.

Individual/Crew Trainers

M1A1/M1A2 Driver Trainers

The tank driver trainers allow trainees to learn and practice combat driving skills in a wide variety of terrain, visibility, and weather conditions. The systems include a high fidelity simulation of the driver's station and the environment, including a 6-degree-of-freedom (DoF) motion system and Compuscene PT2000 image generation system. Three monitors provide a 132° horizontal field of view (Strachan, 1998). The individual driver stations have a fixed configuration representing either the M1A1 or the M1A2 Tank. Four of the M1A2 simulators are being modified to simulate the System Enhancement Program (SEP) upgrade, which includes enhanced digital displays.

The system is used for Armor One Station Unit Training (OSUT). The simulator supports 100 training scenarios that vary according to the location and type of terrain, visibility, time of day, condition of the hatch, and operation in a nuclear, biological, and chemical (NBC) environment. Trainees in OSUT drive about 50 miles on the simulator, while they use actual tanks for about 12-18 miles. The manufacturer (Lockheed Martin Information Systems) states that the reduced cost of driver training – due to reduced fuel consumption, vehicle maintenance and downtime, and avoidance of third-party damage – has led to a cost savings estimated at over $90 million during the first 5 years of operation.

The instructor/operator station allows for control of scenarios and provides trainee feedback. The trainee is given fairly limited performance information, consisting primarily of Go/No Go information on each exercise. Sometimes the instructor will override the automated performance scores (e.g., if the simulator standards are viewed as too stringent). Replay is also possible, but the simulator only saves the last 2 minutes of the exercise. According to the instructor who was interviewed, the replay feature is rarely used.

M1 Conduct of Fire Trainer (COFT)

The COFT trains tank commanders and gunners in a graded set of gunnery exercises. The simulator includes a high fidelity representation of the gunner and Tank Commander (TC) stations in the M1 turret. The representation is accurate except that in the simulator, the turret movement controls are electrical rather than hydraulic. The electrical controls are more sensitive than hydraulic controls, making the simulator somewhat more difficult to control. Later models of COFT use the Compuscene PT-2000 image generation system. COFT is also available in configurations that simulate the M2/M3 Bradley Fighting Vehicle.

Trainees go through a series of over 200 preset exercises, from single targets to multiple targets, from stationary targets to moving targets, from near to far, from good visibility to bad visibility, and under several specific conditions (e.g., wearing NBC mask). The simulator manages the exercises, and increases the difficulty of exercises as trainee ability increases. All tankers use the COFT, first to learning basic gunnery skills, later to maintain these skills. In addition, TCs and gunners who have not been together long use COFT to improve their coordination.

The AGTS is an enhanced version of the COFT that can simulate the M1A2 in addition to the M1A1 and LAV-25. In addition, the AGTS is deployable, configured in either a container or trailer.

Collective Trainers for Ground Operations

Simulation Networking (SIMNET)

The SIMNET system provides a networked simulation environment that gives platoons, companies, and battalions the capability to conduct force-on-force exercises. Based on research sponsored by the Defense Advanced Research Projects Agency (DARPA) and the Army in the mid 1980s, SIMNET was provided to armor and mechanized units during the 1990s. Over 250 SIMNET simulators have been developed and installed at locations both within and outside the continental United States. In addition, mobile sites consisting of four modules have been fielded.

The SIMNET facility at Ft. Knox includes 41 M1 modules (eight more could be hooked up), and 14 M2/M3 modules. The limited controls and displays of these modules are focused on the maneuver and engagement tasks that are simulated in SIMNET. The specific capabilities and limitations of the modules are summarized by the following points.

- There is an eight-channel visual system (for the M1: three driver vision blocks, three TC vision blocks, a gunner's sight, and a loader's vision block). Although the tank commander has a 360-degree vision through vision blocks, there is no simulation of hatch open operation.

- The only weapon system that is simulated is the main gun. The main gun is bore sighted and zeroed.

- The participants wear a headset with microphone, rather than a helmet.

- Movement is restricted by water features; that is, fording is not possible.

- Loading the main gun is simulated by pressing buttons; there is no actual handling of ammunition.

- Refueling and repair is possible, but is relatively unnatural. After the TC calls for repairs, the simulation waits an appropriate amount of time. Then a repair or refueling truck appears next to the disabled tank. Repair time is realistic, depending on the problem with the system.

- The sound system reproduces track noises, turret movement, weapon firing and battlefield noises.

- A seat shaker simulates the vibration of the engine.

- Networked citizen's band (CB) radio hardware is used for all communication.

SIMNET provides substantial capabilities to provide trainees feedback regarding their performance during an exercise. The feedback capabilities of the system include the ability to display the actions conducted in the exercise on the simulated terrain from any selected viewpoint. This capability can provide useful information during an After-Action Review (AAR). For example, in an AAR that we observed, which concerned an exercise conducted as part of the Officers Basic Course, portions of the exercise were replayed to underscore some of the points made in the AAR. Specifically, at one point, one of the tanks was not in a position to observe approaching enemy. At another point, another tank could not fire because another tank from that platoon was in the way.

Close Combat Tactical Trainer (CCTT)

The CCTT provides enhanced capability for conducting force-on-force exercises through the company level. The CCTT configuration at Ft. Knox can train five platoons or two companies simultaneously. However, current staffing does not allow control of five simultaneous exercises. They have done battalion exercises in which some of the units were represented using semi-automated forces (SAF), but battalion exercises are not part of the system design. The system currently consists of 14 tank modules, 11 Bradleys, 1 FIST-V, 2 dismounted infantry and 2 HMVs. Additional modules were desired to bring the total to 44 tanks and 16 Bradleys.

CCTT is currently located at four CONUS sites, with plans for additional sites. At Ft. Knox, the CCTT is used for the basic and advanced officer courses, and for 19K and 19D Advanced Noncommissioned Officer Course (ANCOC). It is also used for Reserve Component (RC) training on weekends (although funding for this activity has been reduced), for training both Active and Reserve Component Marine units, and for units in the Canadian Army.

The CCTT is a networked simulator that represents a substantial enhancement of the capabilities of SIMNET. Each module uses an Evans and Sutherland ESIG 4530 as its image generation system. Individual modules are a much closer and complete representation of the actual equipment. The enhanced technology of CCTT provides the following capabilities that are not included in SIMNET.

- Modules look like the real vehicles.

- Most of the functionality of the equipment is simulated in CCTT.

- The commander's hatch works. The commander can open the hatch and look around. The simulator tracks the commander's head position and displays an appropriate portion of the field of view.

- There is a 24-hour clock, and the environment changes based on the time of day. Shadows move, the sun and moon are in the appropriate location in the sky, and the light level changes.

- The light conditions can also be manipulated. The cloud ceiling can be lowered, and fog can be introduced into the visual scene.

- The range of view is unlimited, and is constrained only by the terrain and the atmospheric conditions.

- CCTT has night vision capability, including the gunner's thermal imaging site (TIS), and simulated night vision goggles for the commander.

- The representation of vehicles is much more detailed. Vehicles can be identified visually, rather than relying on color coding or bumper numbers, as was done in SIMNET.

- Vehicle weapon capabilities are more realistic.

- All weapons are represented except the loader's machine gun.

- Reloading and refueling takes a realistic amount of time

- It's possible to modify the terrain, for example, to dig firing positions.

Despite these additional capabilities, there were several areas in which system managers identified needs for further enhancements, including the following:

- Plows are only available on SAF modules, not on manned modules.

- There is a need to be able to override the SAF module to force red forces to perform desired activities.

- There is a need for some bumper marking and battle board graphics.

- It should be possible to tailor unit numbers to correspond to the numbers of unit being trained.

One problem mentioned that is not directly related to specific technical capabilities of the system is that CCTT training is often not incorporated into unit training strategies. This problem may occur, in part, because it is difficult for units to develop training tailored to their specific needs. Their CCTT support team can develop training much more easily because of their familiarity with the system.

Collective Aviation Trainers

We had an opportunity to observe an aviation training exercise (ATX) that was being conducted to train an Army aviation brigade preparing to deploy to Bosnia. The ATX focused on the command and staff elements from company to brigade level. The exercise scenario was based on the mission that would actually be performed when the brigade was deployed and on the training needs perceived by the brigade commander. This simulated exercise was followed by a live Mision Readiness Exercise (MRE) at the Joint Readiness Training Center (JRTC).

Brigade, battalion, and company level TOCs were simulated. In addition, pilots flew simulated missions using the Fully Reconfigurable Experimental Devices (FREDs) in the Aviation Testbed, as well as the Combined Aviation Virtual Trainer (CAV-T). The exercise also included live elements representing foreign officials, news crews, and so forth. Mission planning was conducted using the operational systems that would be employed in Bosnia, namely Falcon View and TOPSCENE.

Aviation Test Bed

The testbed was implemented in approximately 1990, although there have been several upgrades to the capabilities since then, particularly in the area of graphics. The testbed consists of seven networked FREDs, one High-Mobility Multipurpose Wheeled Vehicle (HMMWV) simulator, a fixed wing simulator, two stealth terminals, and controller stations. The system is well-used, and usually runs two shifts each day. It is used for the following activities:

- ATXs for units that will be deploying to Bosnia
- Officer Basic Course (Initial Entry Rotary Wing [IERW])
- Warrant Officer Basic Course (IERW)
- Officer Advanced Course (to train staff level battle planning)
- Training of National Guard and Reserve elements (for upgrade and transition training).
- Familiarization training for all users (takes about 2 hours).

The FREDs are reconfigurable and can represent AH-64A, AH-64D, UH-60, CH-47, AH-1S, UH-1H, or OH-58D aircraft. The stations have three seats to accommodate both side-by-side and front-and-back configurations. The fixed wing station can represent an A-10 or an F-16. Overall, only a small portion of the cockpit is represented in the system – much along the lines of SIMNET. The visual display is more detailed than SIMNET, but not as detailed as CCTT. Because the testbed was developed as a research system and upgraded several times, there is not any documentation regarding its capabilities.

CAV-T

The CAV-T is the proof-of-principle prototype of the AV-CATT. Stations are reconfigurable and include a higher fidelity representation of the cockpit than the FREDs.

Cockpit controls are represented on touch-screen displays. Removable panels are used to block out portions of the display that would not show on the actual aircraft. The CAV-T has a more complete representation of the cockpit displays and controls. The stations have two seats next to each other. A panel can divide these seats, when the station is set to represent an aircraft such as the AH-64, in which the pilot sits behind the gunner. The CAV-T can represent the AH-64A, the UH-60A, and the OH-58D.

Out-of-cockpit visuals are shown through a helmet-mounted display (HMD). A head tracker monitors the position of the head and presents the appropriate visual display. The display corresponds to the field of view of the particular aircraft. For example, a chin window view will be represented if the simulated aircraft has one.

Because the CAV-T is a more accurate simulation of specific aircraft, it requires a qualified crew to operate. On the other hand, the FREDs can provide meaningful training for students in IERW because they are easier to use.

Mission Planning Equipment Used in Collective Training

Falcon View

Falcon View is a mission-planning tool that is currently used by the Air Force. This is the tool that the soldiers deployed to Bosnia use when they are deployed. The tool consists of a digital map over which the pilot can overlay planned route, waypoints, enemy positions, and other features. Falcon view provides a description of the route that can be used by Top Scene to rehearse the mission.

TOPSCENE

This system provides a high-resolution three-dimensional display of the planned flight. The display system is based on photographic imagery (and other sources); the unclassified version that was used in the exercise had a resolution of 1 meter^2. The operational version has even better resolution. Because of the accuracy and resolution of the display, the pilot can identify visual cues representing waypoints or enemy locations. The system also displays the threat envelope indicating the acquisition and kill ranges of threat targets. This system was originally fielded by the Navy, but has been used by other Services as well. It is currently used in Bosnia. The estimated cost for one such system is $400K (desktop version).

Army Mission Planning System (AMPS)

AMPS is the Army's mission planning system, and, as such, is the analog of Falcon View. This system was not used in the exercise, because it is not being used to its full extent in Bosnia. The major capability of AMPS that is not included in Falcon View is the ability to directly input flight plan information into the helicopter. AMPS produces a cartridge containing the flight plan, which is then inserted into the helicopter. Since AMPS is not compatible with Falcon View, the pilots must manually input the flight plan developed in Falcon View into AMPS.