IMPROVING SOLDIER FACTORS IN HUMAN PERFORMANCE MODELS

Rick Archer, Brett Walters, and Alia Oster
Micro Analysis & Design, Inc
Boulder, Colorado

Dr. Stephen L. Goldberg
U.S. Army Research Institute
Orlando, Florida

Abstract

This paper describes work being performed under Phase II of a Small Business Innovative Research Project (SBIR) for the Army Research Institute. A goal of this effort is to improve the realism of computer-generated force CGF entities in constructive simulations. Currently, CGF entity behavior is very predictable and unrealistic with respect to the natural variability with which humans perform given varying amounts of aptitude, training, and environmental stressors to which they would be exposed on a real battlefield.

A second goal of the SBIR effort is to develop a set of algorithms and data structures for including variables such as aptitude, training, and stressor effects that can be integrated with other types of available software packages used for developing human performance models.

The work is currently progressing on three main thrusts. One thrust is focused on developing the aptitude algorithms, learning curves, and stressor algorithms that will eventually influence the performance variables in the human performance models. Work in support of this thrust included a review of the literature on individual and team learning theories, military requirements for human performance modeling and a search of the literature for empirical data that describes the effects of training on human performance. Work toward developing the learning curves has also included the development of a data collection questionnaire for obtaining estimates from soldiers on how their training affected their proficiency in combat performance. This questionnaire was administered to platoon leaders and platoon sergeants from Armor divisions at Fort Riley, Kansas and Fort Carson Colorado. Data from the questionnaires has been analyzed and used to develop learning curves for classroom, simulator, and field training effects.

A second thrust of the project is the development of a software tool that will allow a user to enter information about the training and aptitude of a population of operators in a human performance model. This information will be used as input to the learning curves to calculate the appropriate changes to performance variables in the human performance models. In order to make the tool generalizable for any type of training or human performance model, we are designing it to be completely configurable by the user. This feature of the software tool will allow users that have data to generate their own learning curves use those learning curve algorithms to affect the performance variables in the models. A part of this effort was the development of a test bed model that we are using to test and demonstrate the functionality of the software tool in correctly modifying performance variables.

The third major thrust of the project is to develop the architecture for communicating performance variable values between the software tool and the human performance models. Included in this effort is the selection of an appropriate entity-based constructive simulation such as ModSAF, OneSAF testbed, or JointSAF to apply the tool described above. At this time, the eventual target platform is the OneSAF test bed. However, since all three of these simulations share the same code base, we have begun this work on the JointSAF software. One of the major challenges in developing the communication architecture is in identifying the performance and potential stressor variables in the simulation that can be modified.

This paper presents details of the work that has been and work that is yet to be done for each of these three major thrusts.

Biographical Sketch

Rick Archer is a Principal Analyst at Micro Analysis and Design. He was involved with the early design and development of the Micro Saint simulation software and has been developing simulation models for 15 years for a number of training, human performance, and process reengineering applications.

Dr. Stephen L. Goldberg, Simulator Systems Research Unit, U.S. Army Research Institute. Dr. Goldberg has been Chief of the ARI Simulator Systems Research Unit for twelve years. He has been active in training research for 26 years. His interests include behavioral modeling, dismounted soldier simulation and training effectiveness.

Brett Walters is a Human Factors Engineer at Micro Analysis and Design in Boulder, Colorado. He received a B.S. in Computer Science and Psychology from Allegheny College in 1995 and a M.S. in Human Factors Psychology from Wright State University in 1997.

Alia Oster has been an Analyst at Micro Analysis & Design for the past two years. She has a Bachelor’s degree in Applied Mathematics with a minor in Computer Science from the University of Colorado. Her responsibilities on this project include algorithm and software development.

IMPROVING SOLDIER FACTORS IN HUMAN PERFORMANCE MODELS

Rick Archer, Brett Walters, and Alia Oster
Micro Analysis & Design, Inc
Boulder, CO 80301

Dr. Stephen L. Goldberg
U. S. Army Research Institute
Orlando, FL 32826-3276

INTRODUCTION

A key consideration made by decision makers at the highest levels of the U.S. military is the trade-off between the acquisition of new weapon technology and allocation of budget resources for training. Researchers at the Institute for Defense Analyses have conducted computer simulation experiments of scenarios from the Persian Gulf War that build a powerful argument that without a major skill advantage held by the U.S. troops, the outcome of the war could have been radically different in spite of the superior technology (Biddle, Hinkle, and Fischerkeller, 1999).

To address the skill acquisition and readiness concern, military leaders are looking more and more to Advanced Distributed Simulation (ADS) and Distributed Interactive Simulation (DIS) as an alternative to highly expensive field training. Using ADS or DIS, soldiers in geographically separate locations can participate in training exercises and perform as if they were on the same battlefield. However, due to the cost of full motion simulators and other logistical considerations, some of the participants in the simulated battlefield exercises are actually computer generated forces (CGF). Human participants in the combat simulation exercises can look out of their vehicles through windows, sites, or whatever means is available and see other participants in their vehicles or combat systems. Some of these other participants are other soldiers in simulators. Others are computer generated.

The problem with the CGF is that they don't behave in a realistic manner. CGF in combat simulation exercises behave in accordance with the engineering specifications of the vehicles and weapon systems that they operate and in compliance with military doctrine but do not reflect the variance of realistic human performance. The performance of the CGF doesn't reflect different amounts and types of training they have received, the experience level of the operators, or the levels of environmental stress such as fatigue, heat, noise, or the use of protective clothing.

The lack of realism in CGF creates some significant shortcomings in the effectiveness of the combat simulations. The first is that the human soldiers can usually tell which of the vehicles and weapon systems on the simulated battlefield are being operated by other humans and which are CGF. This allows them to "game" the simulation exercise and essentially behave differently than they would if they believed that other humans were operating the other forces. Another shortcoming is the lack of the ability of soldiers to train against computer generated enemies with differing levels of training, experience, and/or stress levels.

To address these issues, a Small Business Innovative Research (SBIR) project was funded for the Army Research Institute to develop the capability to manipulate human performance variables that could affect the behavior of CGF. The goal of the effort is to predict the combined effects of training, aptitude, experience and environmental stressors on the performance of simulated humans in CGF.

Although still in progress, the work has proceeded along three major thrusts. First is the development of algorithms that are used to manipulate the human performance variables in constructive simulations that contain CGF. The second thrust is the development of software tools that allow a researcher or combat modeler to enter data that reflect the parameters of the variables (e.g. how much training, aptitude level, etc.) that are used by the algorithms to affect performance. The final thrust is the development of an interoperability architecture to communicate the performance variables to the CGF entities in the constructive simulations.

ALGORITHM DEVELOPMENT

To meet the goals of the project, several different algorithms needed to be either developed or found in the literature. These included a training effects algorithm, algorithms for calculating the effects of aptitude and experience on performance, environmental stressor algorithms and a methodology for combining the effects of each performance variable. The algorithm development began with an extensive review of the literature on each of the performance modifiers.

Training Effects

The review of the literature found that research on the effects of training on performance was extensive but not quantitative in terms of the absolute performance effects derived from a specified amount of various types of training. The quantitative data that were found in the literature was typically very task specific. For example, a report examining the feasibility of developing a stand alone, quantitative Human Reliability Growth Model (HRGM) that predicts the impact of training on soldier performance (Lowery, Rappold, and Copenhaver, 1992) concluded that, although a theoretical basis for developing an HRGM exists, the data could not support its development.

The objective of this SBIR project is to develop training effects algorithms that could be generalized to a wide variety of military tasks that are commonly found in ADS or DIS models. In lieu of generalizable empirical data on training effects from the literature, it was decided to collect data directly from soldiers about the training they received and the effect the training had on their performance.

The project team developed a survey that asked questions about each of the 38 top level tasks from the ARTEP 17-237-10 Mission Training Plan (MTP) for Tank Platoons. The questions pertained to starting proficiency, amounts of classroom, simulator, and field training, and ending proficiency for the previous year. The survey was administered by interview with Platoon Leaders and Platoon Sergeants from Armor Divisions at Fort Riley, Kansas and Fort Carson, Colorado.

After the questionnaire data were collected from the soldiers, a mathematical methodology was developed to combine the data and make it generalizable to other military tasks. The first step was to develop a training taxonomy. The taxonomy was designed to reflect the kind of task groupings on which soldiers are trained. The taxonomic categories that were developed are:

- Planning
- Communication
- Command and Control
- Technical proficiency with equipment
- Tactics and Doctrine

Next, the MTP tasks were mapped to the taxon categories. Each MTP task could be mapped into more than one taxon category. For example, the task 'Conduct troop leading procedures' might be partially in the planning taxon, partly in the communication taxon, and partly in the command and control taxon. Figure 1 illustrates the task to taxon mapping.

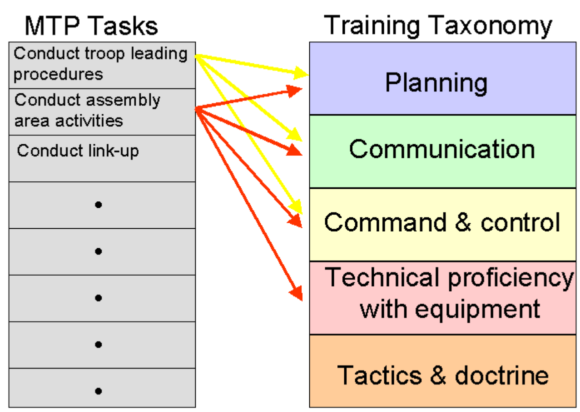

Figure 1. Task to taxon mapping

Another focus of the literature review was to study the current skill acquisition theories. It was found that there are four basic learning curves that are commonly accepted in the training and skill acquisition community. In a review of nearly 3,000 research titles and abstracts, Lowery, Rappold, and Copenhaver (1992) identified these curves as the power law curve, the exponential curve, the hyperbolic curve, and the logistic curve. The power law curve was selected as the curve to be used for the training effects algorithm. Learning rates for the curve were derived from the soldier survey data. To properly model training effects, a learning curve was created for every combination of taxon and training type. For a detailed discussion of the development of the learning curved algorithms see Oster, Walters, Archer, and Edgin, (2001).

In addition to learning curves for skill acquisition, a skill decay curve algorithm was developed. It has been found that different types of skills decay at different rates without additional training or practice. The decay curves that were developed are a variation of the power law curves used for skill acquisition. A decay curve was developed for each of the training taxons. Data for the decay rates were also derived from the soldier surveys of MTP tasks.

Stressor Effects

Other variables that can be used to make entity behaviors in CGF models more realistic are physiological stressors. One of our goals for Phase II of this project was to integrate stressors with the training effects algorithm. Performance on a task or mission under ideal conditions may differ drastically from performance on the same task or mission under stressful conditions. Since the environment in which military operations (tasks and missions) are conducted can be very stressful, incorporating stressors into simulation models allows users to estimate mission performance under realistic conditions.

A review of the relevant literature has produced performance degradation factors for the following stressors: heat, cold, noise, fatigue, circadian rhythm, Mission Oriented Protective Posture (MOPP) gear, and altitude (Bradley & Robertson, 1998; Micro Analysis and Design Incorporated & Dynamics Research Corporation, 1999; Walters & French, 2000). Most of these degradation factors have been correlated to affect a specific set of taxons (e.g., visual, numerical, and cognitive skills). Fortunately these taxons are generalizable and can easily be related to a wide variety of tasks.

Empirical and theoretical research has also been conducted on the combined effects of multiple stressors and how to model them. Although an abundant amount of empirical data has been collected on the separate effects of stressors on performance, there has only been a small amount of work done on the *combined* effects of *some* stressors. Therefore, difficulties occur when equations are developed that try to generalize the interactive effects of multiple stressors. Two different human performance modeling tools were identified that contain equations that address the combined effects of stressors: the Integrated Performance Modeling Environment (IPME) (Micro Analysis and Design Incorporated, 1999) and the Improved Performance Research Integration Tool (IMPRINT). Both of these equations are reasonable for modeling many interactions amongst stressors. However, they may not be reasonable for all interactions. Note that this work is not dependent on the IPME or IMPRINT tools for functionality. Rather, they are sources for equations that attempt to explain the complex relationship between multiple stressors and performance. As more empirical data become available, these equations will be re-evaluated. Work on combining stressors is currently being studied by the UK DERA Center for Human Sciences. They are collecting human performance data under various combinations of stressors specifically to explore the appropriate structure and mathematics that should be used to model combined stressor effects.

For now, we have decided to use an equation found in the literature (Harris, 1985) to combine the effects of multiple stressors. Figure 2 illustrates the conceptual description of the stressor algorithm.

Using this equation, when two or more stressors are combined, the overall degradation is *less* than the sum of the individual degradations. The most severe stressor will have a full effect on performance. As additional stressors are added, they will have less and less impact on performance.

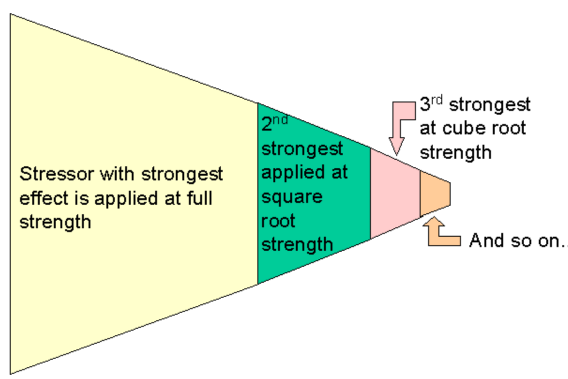

Figure 2. Combining stressor effects

In order to generalize the effects of the physiological stressors to a wide variety of tasks, it was also necessary to develop taxonomic categories for the tasks. The taxonomy that is used describes the composition of each task in terms of perceptual, cognitive, motor, and communication requirements.

Aptitude

The review of the literature also revealed, not surprisingly, that aptitude has a significant effect on performance, especially as it relates to the level of performance that can be achieved with a specified level of training. For example, a 1969 study done for the US Army found that individuals in the bottom 20% of their ability distribution required up to five times as much instruction and practice to attain minimal proficiency in basic military tasks such as rifle assembly (Gottfredson, 1997; Sticht, Armstrong, Hickey, & Caylor, 1987).

General ability has also been shown to predict job performance, and in more complex jobs it does so better than any other single personal trait, including education and experience (Gottfredson, 1997). For example, the Army's Project A, a study conducted in the 1980s to improve the recruitment and training process, found that general mental abilities were highly correlated with technical proficiency and soldiering proficiency (Mchenry, Hough, Toquam, Hanson, & Ashworth, 1990).

For this project, the effects of aptitude were calculated for task times and decision accuracy using data from Project A. An important step in our calculations is to assign an entity's aptitude level. A general aptitude factor can be extracted from skill batteries such as the Armed Forces Vocational Aptitude Battery (ASVAB), intelligence category (CAT) and standard intelligence tests (e.g., Gottfredson, 1997). In this project, it can be done in one of two ways: 1) selecting the Military Occupational Specialty (MOS) of the entity (soldier) performing the task or 2) selecting the minimum ASVAB composite score of the soldier that can perform the task. Each MOS is assigned a composite(s), which is best for measuring the skills and aptitudes required to perform jobs well in the MOS.

In our algorithms, each task is weighted based on how much of the following taxons (skills) are required to perform the task: visual, numerical, cognitive, fine motor discrete, fine motor continuous, gross motor light, gross motor heavy, communications (read/write), and communications (oral) (This is because the effects of aptitude vary depending on the taxon). Then, the aptitude of the entity being modeled is compared to the aptitude for the MOS 12B, which is the designation for a tank Platoon Leader (from which our original data were obtained). Based on this difference and the task-to-skill weightings, a multiplication factor is pulled from a lookup table (generated from Project A data) and used to change the task time or decision accuracy.

To summarize the discussion on the development of algorithms that model how aptitude, training, and stressors affect human performance of tasks, common sense tells us that these algorithms be applied in the order illustrated in figure 3.

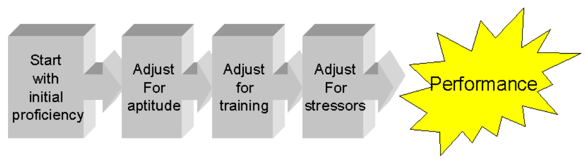

Figure 3. Performance effects

A person comes to a task with some starting proficiency level. Then we adjust the rate at which they can learn to perform the task with an aptitude level. Then we give them an amount of training that changes their level of proficiency as a function of their aptitude and the amount of training they have received. Finally we adjust their proficiency as a function of the current physiological stressors that they are experiencing. The final proficiency level determines the level of performance in terms of speed or accuracy.

SOFTWARE DEVELOPMENT

The second major thrust of this SBIR effort is in the development of software that will allow a researcher or combat modeler to use the algorithms that were developed to actually change the behavior of CGF in constructive simulations. The software is being developed as two distinct modules. One module is basically a user interface that is embedded in a constructive simulation. The simulation that was selected as the first implementation for this work is the OneSAF Testbed Baseline (OTB) model developed by the Army. The purpose of this interface is to allow a OneSAF Testbed user to specify information about the aptitude, amount of training, and stressor parameters for each simulation entity in the combat exercise.

When the OTB users choose to enter aptitude, they are presented with the default MOS and ASVAB cut-off score. The user can then either select a different MOS which will display a new (higher or lower) cut-off score or select a new cut-off score. When the user accepts the current parameters, those are the values that are used to calculate the aptitude of the entity.

When a user chooses to enter training parameters, the user interface displays input fields for each training type and taxon to enter the initial proficiency of the entity, the number of hours of training it has received, and the number of weeks since that training occurred.

When users select the ‘Stressors’ option from the OTB interface, a list of the available stressors displays. Each stressor has a check box associated with it. Users check the stressors that they want to be in effect during the simulation. This window also displays a ‘parameters’ option. When a stressor in the list is highlighted, the ‘parameters’ option displays a window showing default parameters for that stressor (e.g. temperature level for the heat stressor). Users will also have the option to ‘subscribe’ to variables in the environmental model portion of OTB, which will be automatically updated as the OTB simulation proceeds through execution.

When the user executes a OTB exercise and has chosen to use the aptitude, training, and/or stressor functionality, parameters entered from this interface are sent via HLA to the performance server (described later in this paper). The performance server contains a performance calculator that uses the various algorithms described earlier to modify entity behaviors in the OTB exercise.

The second software module is the performance server and performance calculator. The performance calculator uses input from the OTB user to calculate performance values (e.g., task level time and accuracy) for the OTB simulation exercise.

However, in addition to just using the algorithms described earlier to calculate the performance values, the software module also contains an interface through which the user can enter his or her own training or stressor data. For example, the data that are currently in the performance calculator uses the power law learning curve and data on learning rates derived from the soldier surveys. It may be that a researcher may someday obtain or conduct experiments that result in better or different data on the effects of learning on performance. To accommodate this, the software allows users to enter their own training types and taxonomies and/or use their own data for learning and decay rates that plug into the curves. In the future, it is hoped that users will also be able to enter their own equations for learning and decay curves.

With the exception of the fatigue stressor function, the other stressor data are in the form of look-up tables that contain performance multipliers. These performance multipliers were derived from empirical data in the literature. However, due to lack of availability, these data are not complete. To accommodate the acquisition of better or more complete data sets, the values in all of the look-up tables are modifiable.

All of the values for all of the algorithms are stored in a database as a library Performance Exercise Set. Users enter their own data and performance parameters by copying this library set and saving it as a new user Performance Exercise Set. Users can create as many sets as they wish. The library set itself cannot be modified.

COMMUNICATION ARCHITECTURE

Performance Server

There are two communication architectures that are currently under development. The first is when the interfaces are incorporated into the constructive simulation (OTB). The interfaces are where the user enters in the amount of training, aptitude, or stressor values that an entity has had will be incorporated into the constructive simulation. The actual training algorithms and human performance models will be separate. This means that the three separate entities, the constructive simulation, the human performance calculator, and the human performance server will need to be able to communicate with each other.

The architecture between the human performance server, the human performance models, and constructive simulation is based on previous performance work done for ModSAF by Micro Analysis and Design. The relationship between the performance server and constructive simulation is a client-server architecture. This architecture is a federation of disparate simulations. These simulations are linked together to accomplish an overall objective of providing an architecture that allows training and various stressors to impact entity behaviors and performance. Figure 4 shows the client-server architecture; the modules within the dotted box are part of the human performance system.

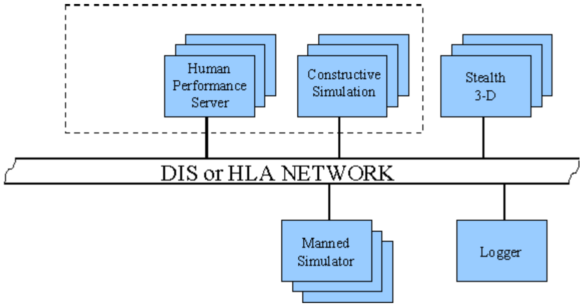

Figure 4. The client-server architecture

The human performance server determines the constructive simulation entity behaviors and performance levels based on the algorithms contained in the human performance calculator. The relationship between the human performance server and the calculator is detailed in Figure 5. The user will enter in the amounts and types of training experienced by the soldiers via an interface in the constructive simulation. These training data will be sent to the performance calculator via the performance server. The calculator will determine the effects the training and stressors will have on performance and will send this information back to the performance server. This way the performance server will be able to provide the constructive simulation performance variable values (i.e. task time and accuracy) during the course of a simulation.

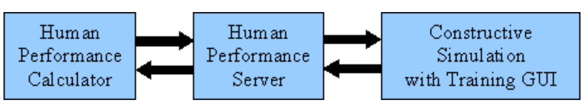

Figure 5. Communication architecture when tool is incorporated into the constructive simulation

The second communication architecture the project team is working on is making the training interface its own software tool, separate from the constructive simulation. This tool will be used with human performance models developed with other modeling tools, such as Micro Saint. Figure 6 shows the architecture for this situation. A user will build a model in Micro Saint and the performance effects software tool will be able to communicate with the model. This allows the tool to see what tasks are in the model. Once the user has set up his or her taxonomy and stressors in the performance tool he will be able to run the model. The tool will tell the human performance calculator how much and what type of training the operator has had. The human performance calculator will use this information to calculate performance values. The model will request these performance values while the model executes.

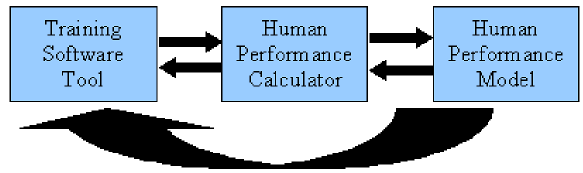

Figure 6. Communication architecture with general training tool

Identifying Performance and Stressor Variables

In order to develop the human performance server we must determine what performance and stressor variables we can affect in the OTB constructive simulation. Determining the performance variables requires studying the OTB code. We have begun to do this with help from another MA&D team who has already affected task performance in ModSAF 5.0 and who are currently working on affecting task performance in JointSAF. As part of this research we are also looking into which environmental stressors, such as heat and humidity, are currently monitored in OTB. Once we have determined the stressor variables that are presently available to us we will work on incorporating them into our performance server and the performance calculator. As part of this investigation we will also determine how we want to let the user incorporate stressors that are not already controlled by the simulation.

SUMMARY

While this project is not complete, it has provided two advances of interest to the training and human performance modeling communities. First, we have successfully developed a method for developing quantitative links between training and performance. Second, we have collected empirical data from soldiers to anchor and populate the resulting algorithm. We believe that these advances will help the Human Systems Integration community impact system, organization, and training system design.

REFERENCES

Biddle, Hinkle, & Fischerkeller (1999, Summer). *Skill and Technology in Modern Warfare. Joint Force Quarterly, 22, 18-27.*

Bradley, C. M., & Robertson, K. A. (1998). *Combined stressors and performance: A review*. Hampshire, UK: Defence Evaluation and Research Agency.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24, 79 - 133.

Harris, D. W. (1985). *A degradation analysis methodology for maintenance tasks*. (Report No. DAPC-OPA-E). Alexandria, VA: Army Military Personnel Center.

Lowery, Rappold, & Copenhaver. (1992). *Feasibility study for predicting human reliability growth through training and practice.* (Report No. ARI-RN-92-39). Alexandria, VA: Allen Corp. of America.

McHenry, J.J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, Special Issue on the Army Selection and Classification Project (Project A)*, *43* (2), 335-355.

Micro Analysis and Design Incorporated (1999). [IPME specification information]. Boulder, CO: Dave Dahn. Unpublished specification.

Micro Analysis and Design Incorporated, & Dynamics Research Corporation (1999). *Enhanced performance degradation factors and upgrades for Improved Performance Research Integration Tool (IMPRINT) Version 5: Stressor review report* (Contract DAAL01-95-C-0122). Boulder, CO: Susan Archer, Rich Adkins, & Brett Walters.

Oster, Walters, Archer, & Edgin. *Mathematical Algorithms for Training Effects Determination if CGF.* Proceedings of the Interservice/Industry Training, Simulation and Education Conference. Orlando, FL. 2001

Sticht, T. G., Armstrong, W. B., Hickey, D. T., & Caylor, J. S. (1987). *Cast-off youth: Policy and training methods from the military experience*. New York: Praeger.

Walters, B., French, J., & Barnes, M. J. (2000). Modeling the effects of crew size and crew fatigue on the control of Tactical Unmanned Aerial Vehicles (TUAVs). In J. A. Joines, R. R. Barton, K. Kang, & P. A. Fishwick (Eds.), *Proceedings of the 2000 Winter Simulation Conference* (pp. 920-924). Piscataway, NJ: Institute of Electrical and Electronics Engineers.