MEASURING THE EFFECTIVENESS OF SIMULATION-BASED TRAINING

Michael L. McGinnis

Department of Systems Engineering, USMA
West Point, New York 10996

George F. Stone III

National Simulation Center, US Army
Orlando, Florida

ABSTRACT

The relationship between measures of training performance and the achievement of training goals is often difficult to quantify. This continues to be an obstacle to the development of curricula for simulation-based training systems. These include both constructive and virtual computer simulations for training. This paper discusses the development of training effectiveness measures as they relate to the design of instructional, simulation-based training systems featuring automated tools for planning training.

1 INTRODUCTION

Recent advances in computer technology have substantially increased the use of simulation-based training by industry, academia, and government agencies. The users of simulation-based training have found it to be an efficient, effective means of training personnel in certain tasks and sustaining the proficiency of those tasks as well. This is especially true for virtual and computer simulation training systems being developed to train personnel as a substitute for dangerous (risky) and resource intensive real-world training. As the future cost of simulation-based training systems decreases, it is expected that they will become a prominent feature of the workplace. During the past several years, the Army has been researching new training performance measures for simulation-based training systems. Agencies involved in this effort include the National Simulation Center, Fort Leavenworth, Kansas, the US Army Simulation, Training, and Instrumentation Command, Orlando, Florida, and the Operations Research Center at the US Military Academy, West Point, New York.

This paper suggests a methodology for measuring the effectiveness of military simulation-based training. Although simulation-based training involves both individuals and teams, our paper is oriented towards team training. Section 2 briefly summarizes previous work in this area. Section 3 outlines a methodological approach for conducting simulation-based training and measuring its effectiveness. Section 4 presents an elementary, yet illustrative, example of a training effectiveness measure developed for simulation-based training.

2 BACKGROUND

A survey of the literature reveals previous studies attempting to define measures of training effectiveness (MOTE) (see Turnage, Houser, & Hofmann, 1989 and Tannenbaum, Cannon-Bowers, Salas, & Mathieu, 1993). Much research focused on specific quantitative or qualitative performance measures within narrowly defined training domains (Spurgin, Moieni, & Orvis, 1993). Others developed methods for conducting and evaluating simulation-based training. The latter generally addressed the training processes, actions, and procedures followed by trainers in conducting simulation-based training.

One of the first systematic training models developed was the Instructional Systems Design (ISD) model. ISD was used during World War II by the U.S. military to train soldiers in aircraft recognition. Recently, Sloman (1994) synthesized the major components of the ISD model into a high-level framework for evaluating training effectiveness. Figure 1 summarizes the stages of Sloman's model.
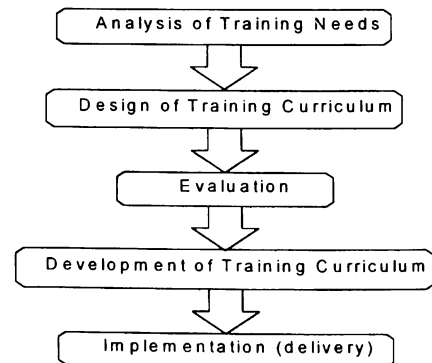

Figure 1. Sloman's ISD systematic training model.

In Sloman's model, training needs derive from a thorough analysis of the training requirements. From this, appropriate and relevant training curricula may be designed and developed. Next, the training curriculum is implemented within a simulation training program. Finally, the training personnel evaluate the proficiency of the team undergoing simulation training. Sloman suggests that the effectiveness of future simulation

training will depend on how well the five components of his ISD model are integrated into training curricula.

According to Williams, Reynolds, Carolan, Anglin & Shrestha (1989), there are several important factors to consider when integrating simulation training into an organization's training program. First, the simulation training must be linked to strategic objectives of the organization. Second, it must also be directly related to the functions and processes of the organization. Lastly, training system development must be perceived by system users as relevant to real-world tasks. In this way training curricula will provide information and training for accomplishing real-world tasks.

3 METHODOLOGY FOR MEASURING THE EFFECTIVENESS OF MILITARY SIMULATION-BASED TRAINING

The approach presented here for conducting and measuring the effectiveness of simulation-based training generally follows the ISD process described above in Section 2. The six major components of our process are: (1) training task identification; (2) training proficiency evaluation; (3) training task prioritization; (4) identification of simulation training support; (5) simulation training execution; and (6) feedback.

Identification of military training tasks for simulation-based training curriculum has benefited from at least twenty years of extensive Army analysis in this area. Previous work determined many tasks for real-world missions that Army units may be assigned. These plans, called *mission training plans* (MTPs), specify the conditions, tasks, subtasks, and standards that Army organizations must meet in order to accomplish combat and peacetime missions. For example, mission training plans may identify the tasks required of an infantry unit to defend key terrain from enemy capture, or a quartermaster unit's tasks for setting up and operating a laundry and shower site for units in the field. Each mission training plan identifies the conditions for performing mission related tasks. They also subdivide major tasks into subtasks with standards for accomplishing each one. Work by the agencies noted in Section 1 included modifying the tasks and subtasks from mission training plans, and other sources, for use in real-world simulation-based training.

The second component is the assessment of a training unit's proficiency at performing mission related training tasks against a measurable standard. It is important that task proficiency be evaluated in a fair, unbiased manner that accurately reflects the training unit's ability to perform key and essential tasks for accomplishing specific missions. Task evaluation may be performed by the team undergoing training or by an outside,

independent evaluator. Recent efforts to automate the task proficiency evaluation process within a Training Exercise Development System (TREDS) are discussed by Crissey, Stone, Briggs, & Mollaghasemi, (1994) and McGinnis & Stone (1995).

In most cases, task proficiency standards are measurable, fixed criteria. Performance outcomes are usually assessed as pass or fail (i.e., go or no go). This helps remove subjectivity and ambiguity that might otherwise bias training performance measurements. See Gonzalez & Ingraham (1994) for further discussion of the application of measurable standards to simulation-based training. Next, we aggregate task and subtask ratings into an overall determination (measure) of training proficiency (see Section 4 below). Based on this assessment, the military unit is rated as either fully trained (T), partially trained (P) but needs practice, or untrained (U).

The third component, prioritization of tasks and subtasks, is a preliminary step to identifying and scheduling training support and resources, i.e., the fourth component, for executing simulation-based training as accomplished by the fifth component (see above). Task prioritization also helps training managers select training scenarios that meet the training goals of the military unit(s) undergoing training. The training task rankings, reflecting the value of each training task to mission accomplishment, and training performance scores obtained during simulation training execution are used to determine training proficiency scores.

The unit receives feedback during an after action review of training performance and lessons learned. Training evaluations provide the military unit(s) with quantitative results in the form of an overall score that benchmarks their performance during training. Training results are also used for adjusting task prioritization to reflect updated training priorities. This marks the start point for the next iteration of the simulation-based training process.

4 FORMULATION OF AN ELEMENTARY SIMULATION-BASED TRAINING EFFECTIVENESS MEASURE

The effectiveness measure presented here for simulation-based training is partially based upon multiple attribute utility theory and decision making methods. See Clemen (1990), Hwang & Yoon (1981), and Saaty (1980) for discussions of the theory and methods.

As explained above, our methodology requires that tasks be identified in advance of training. Following this, each task is ranked (weighted), relative to the others, by the military unit undergoing training. The

weights reflect the importance of each task to mission accomplishment. Precise measurable standards are best for evaluating the training proficiency of the military unit. Figure 2 depicts a weighted sequence of training tasks.
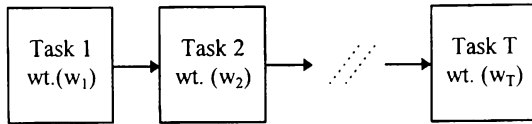

Figure 2. Training tasks with weights

Next, a training task tree is constructed by breaking down each top level task t into lower level subtasks. The highest level tasks, level 1, are usually described in broad, general terms. These high level tasks are the ones that must be successfully completed for the military unit to achieve mission success. At each level, task decomposition leads to more basic, specific subtasks. Working down the tree, lower level subtasks generally reflect "how" the parent task is to be accomplished. Figure 3 illustrates the hierarchical structure of a partial training task tree for a single level 1 task t, decomposed through N subtask levels with L subtasks at each level.
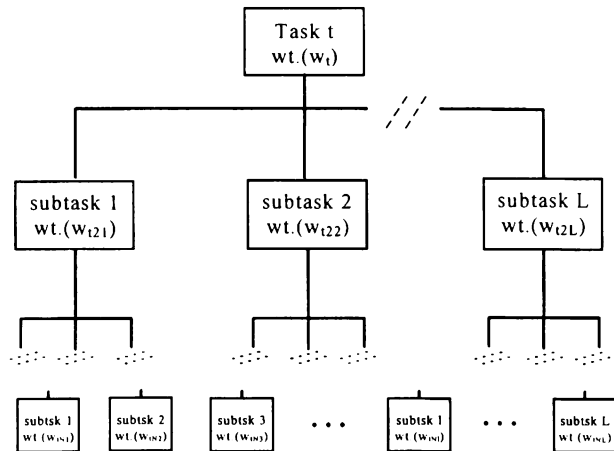

Figure 3. A hierarchical task tree.

Weights must be determined for each subtask. As mentioned above, they represent the importance of subtasks to mission or parent task accomplishment. The combined weights of all subtasks branching from a parent task must sum to one. They may be derived mathematically using various methods or be elicited as subjective assessments from unit personnel.

The tree is solved, branch by branch, starting at the lowest level and working up. A linear combination of weights W and scores S is computed for each branch. These are rolled up to the next branch and added to the

linear combination of weights and scores for the parent node. This process continues for all tasks L, branches M (within a level), and levels N until a total score is computed for the highest level task (level 1). Mathematically, this is given by

$$\sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{j=1}^{L} W_{ijk} S_{ijk} \, .$$

Training proficiency score computations for a hypothetical training proficiency task tree may be illustrated using a simple additive procedure. We note, however, real-world situations may demand a more sophisticated, robust method. Here, we compute the overall task proficiency score for a level 1 task expanded two levels deep. Two branches emanate from the first level to the second level. Three branches emanate from each of the second level nodes.

In practice, task scores may have different units of measurement such as time or distance. Before these task scores can be used to compute a task proficiency score they must be converted into unit-less values called *utiles* (see Clemen, 1990). For this example, we assume all real-world task scores have been converted into utile values as given in Table 1. The utile values, in this case, range between 0 to 10 yielding a utopian value for the level 1 task of 32. Table 1 gives the weights and scores for each subtask, as well as, the combined and aggregated scores computed for each level.

Level	Task	Weight	Task Score	Combined Score	Aggregate Scores
1					**18.2**
2	1	0.4	7	2.8 (+4.8)=	7.6
	2	0.6	8	4.8 (+5.8)=	10.6
3	11	0.3	1		
	12	0.2	10		
	13	0.5	5	4.8	
	21	0.3	8		
	22	0.3	6		
	23	0.4	4	5.8	

Table 1. Training proficiency computation results.

The hypothetical results reveal that the military unit achieved a raw task proficiency score of 18.2 for the level 1 task. This establishes a training proficiency benchmark for this task of approximately 50% of the utopian value.

Training measurement parameters such as task conditions, standards, weights, and utile ranges may vary by training unit, simulation site, or training scenario. Obviously, task proficiency scores are sensitive to changes in these training effectiveness parameters. Therefore, care must be taken to properly

account for such changes when making comparative analyses of a military unit's training proficiency over time using the methodology presented above.

5 CONCLUSIONS

At this time, none of the Army's simulation-based training facilities of Combined Armed Tactical Trainer (CATT) family have yet been built. The first to be developed will be the Close Combat Tactical Trainer (CCTT) for training armor and mechanized forces at battalion and below. The first CCTT facility is currently under construction and is expected to be operational in 1997 or 1998. Therefore, the potential value of the methodology presented in this paper is unknown at this time. Future work will include testing and evaluating the methodology and various training effectiveness measures in simulation-based training experiments to be conducted at a prototype facility located at the US Army Simulation, Training, and Instrumentation Command. Experimentation involving potential future system users will hopefully provide training system developers with feedback for improving both the simulation training systems under development and the methods for measuring training effectiveness.

There is also a high potential for transferring the military simulation-based training technology and training methodology to non-military applications. For example, a System for Training Emergency Personnel (STEP), proposed previously by McGinnis (1994) for training civilian emergency personnel, is very similar to TREDS task identification and the Instructional Systems Design (ISD) model discussed previously. Finally, it is hoped that this work will motivate future research efforts in this area that will lead to improved measures of simulation-based training effectiveness and fielding the best training system possible.

REFERENCES

Clemen, R.T. 1990. *Making Hard Decisions: An Introduction to Decision Analysis.* Belmont, CA: Duxbury Press.

Crissey, M., Stone, G.F., Briggs, D., & Mollaghasemi, M. 1994. Training exercise planning: Leveraging data and technologies. *Proceedings of the 16th Interservice/Industry Training Systems and Education Conference.* Washington, DC: National Security Industrial Association.

Gonzalez, A. & Ingraham, L. 1994. Computer-based exercise progression in simulation-based training. *IEEE Transactions on Systems, Man, and Cybernetics* 24(6): 863-874.

Hwang, C. & Yoon, K. 1981. *Multiple Attribute Decision Making Methods and Applications: A State-of-the-Art Survey.* New York: Springer-Verlag.

McGinnis, M.L. 1994. Unpublished Technical Report: *The system for training emergency personnel (STEP).* The United States Military Academy, West Point, New York.

McGinnis, M.L. and Stone, G.F. 1995. *Recent initiatives to develop decision support systems for command and control (C2) and training planning.* Industry, Engineering, and Management Systems (IEMS) 1995 Proceedings: 584-589.

Saaty, T.L. 1980. *The Analytical Hierarchy Process.* New York: McGraw Hill.

Sloman, M. 1994. *A handbook for training strategy.* Brookfield, Vermont: Gower.

Spurgin, A., Moieni, P., & Orvis, D. 1993. Some thoughts on the use of data for assessing training effectiveness. *Proceedings of the 1993 Simulation Multiconference on the International Simulators Conference.* San Diego, CA: Society for Computer Simulation: 579-584.

Tannenbaum, S., Cannon-Bowers, J., Salas, E., & Mathieu, J. 1993. Technical Report 93-011: *Factors that influence training effectiveness: a conceptual model and longitudinal analysis.* Orlando, FL: Naval Air Warfare Center Training Systems Division.

Turnage, J., Houser, T., & Hofmann, D. 1989. *An assessment of performance methodologies for collective military training.* Orlando, FL: University of Central Florida.

Williams, K., Reynolds, R, Carolan, T., Anglin, P., & Shrestha, L. 1989. Technical report TR89-035: *An evaluation of a methodology for cognitively structuring and adaptively sequencing exercise content for embedded training.* Orlando, FL: University of Central Florida, Institute for Simulation and Training.

AUTHOR BIOGRAPHIES

MICHAEL L. MCGINNIS directs the Operations Research Center at the United States Military Academy (USMA), West Point, New York. He received a B.S. degree from USMA in 1977, M.S. degrees in Applied Mathematics and Operations Research from Rensselaer Polytechnic Institute in 1986, a Ph.D. in Systems and Industrial Engineering from the University of Arizona in 1994, and a M.A. in National Security and Strategic Studies from the Naval War College in 1996. His applied research interests are military command and

control, personnel forecasting, and resource scheduling. E-mail: fm0768@exmail.usma.edu.

GEORGE F. STONE III is United States Military Academy (1980) graduate with a B.S. in General Science. He also received a M.S. in Industrial Engineering from Texas A&M University in 1989, and more recently his doctorate in Industrial Engineering at the University of Central Florida. Major Stone has had numerous field artillery assignments including two battery commands in West Germany. Major Stone is a former Assistant Professor in the Systems Engineering Department at West Point and is currently the Deputy TRADOC Systems Manager for WARSIM 2000. His research areas include simulations, operations research/systems analysis, systems engineering and training.