TEAM PERFORMANCE MEASUREMENT ISSUES IN DIS-BASED TRAINING ENVIRONMENTS

Jennifer E. Fowlkes, Norman E. Lane
Enzian Technology, Orlando, FL

Daniel J. Dwyer, Ruth P. Willis, & Randall Oser
Naval Air Warfare Center Training Systems Division, Orlando, FL

ABSTRACT

Realizing the considerable training potential of Distributed Interactive Simulation (DIS) technology will require training and performance evaluation methodologies. The objective of this paper is to identify key performance measurement issues in DIS-based training environments. The major premise is that, in DIS-based training systems, problems precluding reliable measurement are likely to be aggravated over those encountered in conventional simulation, presenting unusually complex measurement challenges. As a way to describe the inherent problems, a framework is described which a) identifies those factors which adversely affect measurement and feedback in conventional simulators, and b) describes how those factors tend to have greater impact in DIS environments. The identification of these issues is based on consideration of factors known to affect measurement in operational or simulated operational settings. They encompass task sampling, measurement procedures and the nature of trainee differences. Each of the factors is discussed and recommendations are provided for the reduction of their impact.

ABOUT THE AUTHORS

Jennifer E. Fowlkes has been a Research Psychologist with Enzian Technology, Inc. since 1992. She holds a Ph.D. in Experimental Psychology from the University of Georgia. Dr. Fowlkes has over eight years of experience in areas of human factors and training which includes team training and performance, training effectiveness evaluations, simulator sickness research, and performance test battery development. Most recently, her research experience has focused on measuring team performance in distributed training environments.

Norman E. Lane has worked in human factors engineering, training, and personnel systems for more than 28 years. Dr. Lane has managed large scale Navy programs in areas such as simulator evaluation and utilization, training for perceptually altered environments, and performance measurement systems. Since 1987, he has provided consultant services to a variety of government and industry clients. Recent work includes team performance measurement, assessment of performance changes under stress, and research in above real time training.

Daniel J. Dwyer has been a Research Psychologist with the Naval Air Warfare Center Training Systems Division for over 16 years. His MS degree from the University of Central Florida and his Ph.D. degree from the University of South Florida are both in Industrial/Organizational Psychology. His most current research efforts are in the areas of team performance measurement, after action review and debriefing techniques, and distributed training.

Ruth P. Willis received the Ph.D. degree in Industrial/Organizational Psychology from the University of South Florida in 1985. At present, she is a Research Psychologist in the Training Technology Application Branch of the Naval Air Warfare Center Training Systems Division where she has been conducting research concerning the strengths and limitations of a distributed environment for training.

Randall Oser is a Research Psychologist with the Naval Air Warfare Center Training Systems Division. He holds an MS degree in Industrial/Organizational Psychology and is currently pursuing his Ph.D. in Human Factors Psychology at the University of Central Florida. His major research interests include team performance measurement in distributed systems, assessing the impact of automation and glass cockpit technology on crew member performance, and the acquisition of complex skills.

TEAM PERFORMANCE MEASUREMENT ISSUES IN DIS-BASED TRAINING ENVIRONMENTS

Jennifer E. Fowlkes, Norman E. Lane
Enzian Technology, Orlando, FL

Daniel J. Dwyer, Ruth P. Willis, & Randall Oser
Naval Air Warfare Center Training Systems Division, Orlando, FL

Distributed interactive simulation (DIS) is an emerging technology that offers tremendous leverage to the DoD in a number of areas including the development of tactics and doctrine, test and evaluation, and the development of notional weapons systems. A primary anticipated application of DIS is the training of warfighting skills; that is, those skills required to operate platforms as they are integrated with other systems within a coordinated battle or mission scenario (Lane & Alluisi, 1992). DIS has the potential to support unprecedented training opportunities. This, combined with cost savings and safety considerations make DIS a potentially important adjunct to live training. Recent demonstrations include Synthetic Theater of War-Europe (STOW-E) and Prairie Warrior, both of which linked virtual and constructive simulation with live maneuvers on instrumented ranges, and the Multi-service Distributed Training Testbed (MDT2) which provided an environment for the training of multi-service behaviors involved in the planning and execution of close air support.

A key challenge to the advancement of DIS-based training systems is the development of performance measurement methodologies to support training and training effectiveness research. Regardless of the training application, performances in DIS must be observed and evaluated and trainees must be given feedback via after action reviews (AARs) and debriefs. Moreover, as DIS-based training exercises and systems are developed, there will be a concomitant requirement to establish their training effectiveness. Such evidence will be demanded given the costs and time invested in the development of this technology.

In operational or simulated operational settings, successful performance measurement has traditionally been difficult (Lane, 1986; Lane & Kennedy, 1994; Vreuls & Obermayer, 1985). Not surprisingly, in DIS-based training systems, the problems precluding reliable measurement are likely to be aggravated over those encountered in conventional simulation. The objective of this paper is to identify key measurement issues as they affect ability to measure team performance in DIS-based training environments. As a way to describe the inherent problems, a framework is described which a) identifies those factors which adversely affect measurement and feedback in conventional simulators, and b) describes how those factors tend to have greater impact in DIS environments. "Conventional" simulation, as used in this paper, refers to the training of individuals or small teams in a single self-contained training device. Distributed simulation refers to the training of multiple players or teams geographically or physically dispersed across a number of simulation nodes. In general, measurement in DIS environments is influenced both by those factors operative in conventional settings and by those which are specific to or aggravated by the special characteristics of DIS, presenting an unusually complex measurement challenge. Existing DIS technological problems exacerbate the measurement problem.

MEASUREMENT ISSUES

The identification of measurement issues in DIS derives from consideration of factors known to affect measurement in conventional operational or simulated operational settings. These factors include task sampling, measurement procedures, and unwanted variance due to trainee differences. Table 1 provides examples of the differential impact of these factors in conventional and distributed simulation training environments. Each of the factors is discussed in detail below.

Table 1
Factors Affecting Performance Measurement in Conventional and Distributed Simulation Environments

Measurement Factor	Conventional Simulation	Distributed Simulation
Task Sampling		
• Nature of the Task	Generally task content can be adequately specified and sampling adequacy can be determined	Domain of warfighting skills is more difficult to define and, thus, task content is less certain
• Scenario Control	Scenario control can be achieved through imaginative scripting of exercises	Greater number of entities, and their interdependencies, makes scenario control problematic
• Equipment/System Reliability	When equipment goes down, the exercise may have to be terminated but can generally be restarted with little information loss	When a simulation node goes down, the exercise may continue but the nature of the task being performed may change. Restarting off-line nodes requires that those nodes catch up on exercise status.
• Equipment Fidelity--differences between simulators and real world	Resolution inadequacies, field of view limitations, etc. must be taken into consideration for performance measurement	Same as in conventional simulation
• Equipment Fidelity--differences between simulators	Not applicable	Differences in levels of fidelity between different devices results in unlevel playing fields; that is, pairwise constraints exist in detectability/ localization/ identification imposed by equipment mismatches
Measurement Procedures		
• Familiarization/Training of O/Cs	Generally easily accomplished	The O/C preparation task is greater
• Frame of Reference	Generally, individual O/C or instructor inputs are used to support performance assessment	A team of observers is required for performance assessment, each of whom has a different physical viewpoint and possibly a different area of responsibility
• Availability of Required Information	The ability to observe performance is generally not an issue	Global, network-wide observation may be difficult

Table 1 continues

Trainees		
• Differences in Operator Capability	Differences in individual aptitude, training, and experience as a team	The same factors operate as in conventional simulation. In

Table 1
Factors Affecting Performance Measurement in Conventional and Distributed Simulation Environments

Measurement Factor	Conventional Simulation	Distributed Simulation
	member can contribute unwanted variance to performance measures	addition, there is a greater number of participants and participant/ participant interactions.
• Inherent Performance Instability	Day to day performance variability (e.g., due to fatigue, etc.)	Same as above
Participant Understanding of Task Requirements	Understanding is required of own platform tasks, threat characteristics and mission	Understanding is required of own platform plus that of other entities/roles/tasks. Generally there is a heavier planning component. Participants must understand the impact of multiple fidelities on task performance.

Task Sampling

"Task Sampling" concerns the extent to which the tasks that the trainee must perform in the actual simulation are a) representative of those encountered on the job, b) representative of those which the simulation is intended to train, and c) of known content (i.e., we can determine what is being practiced or trained). Factors which affect task sampling include the nature of the task, type and extent of control exercised over the scenario, the nature and extent of equipment intermittency, and equipment fidelity issues; that is, the degree to which fidelities of participating nodes are comparable among the nodes and to the "real world."

Nature of the Task. The development of DIS technology is influenced by the need to provide opportunities to practice "warfighting skills" (those skills which go beyond basic tactics and use of equipment) in multi-service and combined force operations. The task domain defined by warfighting skills will likely be more complex than those task domains encompassed by conventional simulation. The impact on measurement is that the task content measured will be less certain.

Scenario Control. The flow of events in a scenario can be broadly categorized as either "free play" or "scripted." Free play exercises are those for which event flow is largely determined by the give-and-take, real time interactions of players in the exercise. While free play exercises are probably more representative of the actual battlefield, they are extremely difficult to evaluate with respect to training effectiveness, since they allow only product or outcome measures to be determined. The task content in free play exercises is to a large degree left to chance.

Scripted exercises are those for which major events and waypoints are largely governed by preset, realistic constraints and conditions. Such exercises are considerably more useful not only for training specific tasks but also for obtaining evaluation data. If, for example, there is interest in evaluating multi-service interactions in close air support operations, the scenario must provide specific and identifiable opportunities for such interaction events to occur.

Scenario control is more challenging in DIS environments because of the number of alternative actions possible on the part of participants. Lane and Alluisi (1994) noted:

> Battlefield situations are changed dramatically by entity activities, and because of this free play, battle outcomes are never predeterminable, but rather evolve as a direct result of interactions among players in the simulation. (p. 4)

Training utilizing DIS will likely involve the development of detailed battle situations which provide a context for the operations. These battle situations or scenarios are crucial in that they fully determine what one will be able to observe, record, and hence measure in DIS-based training. Such control is important in DIS-based training environments. As Salas, Bowers, and Cannon-Bowers (in press), as well as others (e.g., Modrick, 1986), have argued, it cannot be assumed that training is occurring simply because teams have the opportunity to practice together. Well-designed training will present opportunities for targeted interactions to occur. Although controlled event presentation is difficult to achieve in DIS-based training environments, technologies are being developed for its accomplishment (e.g., Atwood, Winsch, Quinkert, & Heiden, 1994).

Equipment/System Reliability. In conventional simulation, system reliability problems are generally easily recognized by the observer/controller (O/C) or instructor who can account for their impact when evaluating performance. If a system goes down, often the exercise can subsequently be continued with little information loss. In distributed simulation, system unreliability has a more corrupting impact on performance measurement. When a node goes off-line, generally the exercise may continue, but the nature of the task being performed changes. When nodes come back on line, trainees at those sites must catch up on exercise status. Real-time node status information may not be available to O/Cs (Madden, 1994) with the consequence that they may be unable to assess performances accordingly. The picture is further complicated by the degradation of individual simulation subsystems (e.g., navigation systems) at some sites that are likely to go unnoticed by O/Cs at other sites.

There are, in addition, a host of interoperability limitations in DIS-based systems that affect exercise play and the nature of the tasks being performed. Numerous examples could be provided including aircraft flying below the ground and inability to see entities from other sites due to mismatches in terrain data bases. Other examples are provided in Table 2. The point to be made by Table 2 is that these often bizarre effects impact task content. Moreover, they are confusing and disruptive to O/C efforts to measure performance.

Fidelity. In both conventional and distributed simulation systems, a performance measurement issue is the extent to which the simulation system(s) are comparable to the real world system(s). An additional consideration in DIS-based systems is the extent to which fidelities of participating systems are comparable. Altman, Kilby, and Lisle (1994) described the problem:

> Functional and implementation differences between interacting simulators can yield unrealistic advantages for some and deficiencies for others that can adversely affect the training objectives (p. 535).

Fidelity differences among systems is a problem inherent to interoperating dissimilar systems. To achieve meaningful performance measurement, it is important that a level playing field is achieved among interacting entities. If a level playing field cannot be achieved, it is important that trainees and O/Cs fully understand the system fidelity limitations and that these are allowed for in performance measurement. Altman et al. (1994) argue that solutions will be difficult to achieve.

Measurement Procedures

The "Measurement Procedures" factor refers to the methods by which performance is observed and measured. The most critical aspects of measurement procedures are the ways in which observers are trained and used as a measurement source, the opportunities provided for observers to acquire performance-related information, and the extent to which information is lost by intermittency of the network.

Familiarization/Training of O/Cs. O/Cs in DIS-based training, as in conventional simulation, have the dual responsibilities of controlling the training exercise

Table 2

Examples of Interoperability Problems and Impact on Task Execution

Problem	Task Impact	Source
Threat vehicles would appear and disappear	Impacted ability to engage targets with Hellfire missiles	Adams et al., 1994
Invisible solid walls in space which caused aircraft to crash in midair due to disparities in terrain data bases	Global impact on task execution	Adams et al., 1994
Too many target marking rounds fired at one node exceeded entity limitation of another node.	Aircrews unable to visually acquire target	Fowlkes et al., 1994
Loss of visual information to (tank) driver due to limited processing capability	Impaired ability to drive, acquire targets	Leibrecht et al., 1993

and of evaluating trainee performance. The adequacy of their orientation, training and preparation for a DIS training exercise is a critical determinant of effective measurement. It can be argued that the O/C task in DIS-based training environments will be substantially more difficult than in conventional simulation. At a minimum, O/Cs must know what the trainees know, and must have sufficient experience and perspective to evaluate decisions made by participants at key event points and to judge overall level of performance. O/Cs must, in addition, have been made aware of details of scenario content and the specific behaviors to be observed. They must also be thoroughly familiar with the limitations of each node in the simulation so that they may judge performances on a "relative" basis.

Frame of Reference. In distributed environments, a team of O/Cs will be required to adequately observe and judge performance. This has associated with it at least two consequences. First, for effective measurement, more resources will be required over those generally required in conventional simulation. Leibrecht et al. (1993) reported that in a DIS-based training exercise, manned simulators were only partially monitored, adversely impacting the performance measurement effort. Related to this, the O/C workload may be greater. O/C tasks may compete with data collection efforts and a separate set of subject matter experts for performance measurement may be required. Leibrecht et al. (1993) noted,

> The flagging and recording of events by control staff observing a PVD [Plan View Display] did not afford the reliability desired. Competing tasks, distractions, and difficulty in interpreting on-line tactical events resulted in lost and unusable data elements (p. 123).

Second, each O/C will have a different physical viewpoint, training, and responsibility. For example, in the MDT2 effort, different O/Cs were required from each of the services represented (Fowlkes, Lane & Llaneras, 1994). Thus, for performance measurement, the input from multiple observers, who may have different training and backgrounds, will likely be required.

Availability of Information. In conventional simulation, instructional features for a particular device are developed based on a training requirements analysis. However. as individual training systems become interoperable via DIS, the nature of the training requirements change, often without the concomitant revisiting of the existing instructional displays and features. The result is that within-node instructional displays may be inadequate to support performance measurement, especially network wide performance monitoring and feedback. Examples of useful displays include real time information on node status, including whether the node is on or off line and, depending on the training situation, the status of subsystems (e.g., visual or communications systems) (Madden, 1994).

“Making do” with what they have, a related problem for O/Cs occurs when sharing a console or when a display must be used for dual purposes. In these situations, O/Cs may have to compete for the use of displays Adams, Courtright, Farrow and Swicord (1994) noted this problem:

> The need to use SAFOR systems to position and attach the Stealth often precluded the Exercise Control Officer from processing information from units. The Exercise Control Officer therefore had to wait until he could gain access to a SAFOR terminal in order to deal with incoming information from the vehicle commanders. (p. 1)

Trainees

The "Trainees" factor includes the ubiquitous presence of individual differences in ability and experience among individuals and teams in the DIS. Such differences may a) make a particular scenario more or less difficult for some players than for others, and b) allow some players to benefit more from a particular training simulation than others do. In addition, the well-known tendency for performance to vary between and within task repetitions is exacerbated in DIS systems by the greater number of entities interacting as the scenario unfolds. A third factor is that performance of a task under DIS is likely to be more complicated than the same task under conventional simulation; the players must know something about the capabilities and fidelities of the other entities involved and their roles in the scenario. DIS requires a greater emphasis than conventional simulation on pre-mission planning and orientation. The importance of informing trainees of fidelity limitations and other familiarization issues has been observed in several DIS-based training exercises (e.g. Atwood, Winsch, Sawyer, Ford, & Quinkert, 1994; Meade, Lozicki, Leibrecht, Smith, & Myers, 1994; Winsch et al., 1993).

RECOMMENDATIONS

General recommendations for reducing the impact of the factors discussed above are provided in Table 3. The recommendations reinforce the point that obtaining successful measurement requires attention to virtually every aspect of the development and implementation of a training program--from the extent to which trainees understand the task to the extent to which a scenario can be controlled.

Many of the recommendations shown in Table 3 derive from two main threads that permeate the ability to carry out successful team performance measurement in DIS-based training environments. The first is that early and extensive preparation is needed to implement training in DIS environments. This encompasses the development of program objectives and training materials, training and orchestration of O/Cs to facilitate control and measurement, the familiarization of trainees and the dissemination of information to them to ensure that they understand the task. Failure to address each of these areas will have a direct impact on measurement characteristics and quality.

The second is that of system reliability and the general state-of-the-art of the technology available to support training. System reliability and interoperability problems, to the extent that they exist, will be a severely limiting factor since they change the nature of the task, make observation difficult, and result in unpredictable and inconsistent data loss. Some of the technologies needed to support the observation of performances at distributed nodes are not yet mature. Combined, these factors contribute to the "fog of measurement." Recommendations to reduce the impact of these include careful training system design to ensure that the tasks to be performed are reliably supported by the technology.

CONCLUSIONS

The discussions provided in this paper address how measures may be affected by the unique environments provided by the application of DIS technology. Research into how best to use DIS technology to support team/collective training

Table 3
General Performance Measurement Recommendations in DIS-Based Training Environments

Measurement Factor	Recommendation
Task Content	• Define training focus early in training program/system design • Include DIS-knowledgeable engineers on training system design team to ensure that tasks to be trained are reliably supported by technology • Revisit and modify tasks to be trained as interoperability and other engineering tests are conducted • Identify fidelity limitations and ensure these are understood by O/Cs and reflected in the performance measurement system • Develop exercise control contingencies for nodes dropping off line • Develop scenarios that are robust to node intermittency • Utilize scenario control, transparent to trainees, as one means of controlling task content • Ensure explicit links exist between program goals, training objectives, exercise design and performance measurement
Measurement Procedures	• Ensure O/Cs are thoroughly familiar with the goals of the training system, goals for performance measurement, system limitations, etc. • Thoroughly brief O/Cs on the roles and responsibilities of individuals comprising the O/C team and on exercise control mechanisms • In established DIS-based training systems, use a dedicated cadre of O/Cs to minimize the O/C familiarization task • Ensure that observation technologies are present to support observation of performance (e.g., ready and rapid monitoring of node status, communication channel between O/Cs) • Asses likely O/C workload and assign a separate set of observers for the performance measurement task if needed
Trainees	• Brief trainees on the capabilities and limitations present at each site as applicable to their performance of the training tasks • Plan time for trainee familiarization with the simulation; it may be desirable to conduct a network-wide familiarization exercise

is just beginning. Little is known, for example, about how to structure distributed scenarios, how to provide practice, and how to provide feedback. The intent of this effort was to advance future performance measurement efforts in DIS-based training environments.

REFERENCES

Adams, J., Courtright, J., Farrow, S., & Swicord, P. (1994). *Advanced distribution simulation technology Zen Regard experiment - Final report* (Contract Number N61339-91-D-001).

Altman, M., Kilby, M., & Lisle, C. (1994). On a shared environment concept for distributed simulation. *11th DIS Workshop on Standards for the Interoperability of Distributed Simulation, Volume Papers* (pp. 535-543). Orlando, FL: Institute for Simulation and Training.

Atwood, N. K., Winsch, B. J., Sawyer, A. R., Ford, L. A., & Quinkert, K. A. (1994). *Training and soldier-machine interface for the combat vehicle command and control system* (Document Number W003266). Orlando, FL: Loral Systems Company.

Atwood, N. K., Winsch, B. J., Quinkert, K. A., & Heiden, C. A. (1994). *Catalog of training tools for use in distributed interactive simulation (DIS) environments* (ARI Research Product 94-12). Fort Knox Field Unit: U.S. Army Research Institute for the Behavioral and Social Sciences.

Fowlkes, J. E., Lane, N. E., & Llaneras, R. E. (1994). *Performance measurement in distributed interactive simulation (DIS) based training environments: identification of key issues* (Interim Report). Orlando, FL: Enzian Technology, Inc.

Lane, N. E. (1986). *Issues in performance measurement for military aviation with applications to air combat maneuvering* (NTSC-TR-86-008). Orlando, FL: Naval Training Systems Center.

Lane, N. E., & Alluisi, E. A. (1992). *Fidelity and validity in distributed interactive simulation: Questions and answers* (IDA Document D-1066). Alexandria, VA: Institute for Defense Analysis.

Lane, N. E., & Kennedy, R. S. (1994). *Design of visual simulation systems in combined arms tactical training (CATT): Volume II: Resolving issues in training effectiveness evaluation* (TCN 94004). Research Triangle Park, NC: US Army Research Office.

Leibrecht, B. C., Winsch, B. J., Ford, L. A., Sawyer, A. R., Meade, G. A., Ainslie, F. M., Smith, P. G., Sever, R. S., & Doherty, W. J. (1993). *Battalion evaluation of the combat vehicle command and control system in distributed interactive simulation: Preliminary findings* (ARI Technical Report 992). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Madden, J. J. (1994). Exercise control/monitoring issues in distributed close air support exercises. In the *11th DIS Workshop on Standards for the Interoperability of Distributed Simulation, Volume Papers* (pp. 9-10). Orlando, FL: Institute for Simulation and Training.

Meade. G. A., Lozicki, R., Leibrecht, B. C., Smith, P. G., & Myers, W. E. (1993). *The combat vehicle command and control system: combat performance of armor battalions using distributed interactive simulation* (Document Number W003262). Orlando, FL: Loral Systems Company.

Modrick, J. A. (1986). Team performance and training. In J. Zeidner (Ed.), *Human productivity enhancement: Training and human factors design in systems design. Volume I.* Santa Monica: System Development Corporation.

Salas, E., Bowers, C. A., & Cannon-Bowers, J. A. (in press). Military team research: Ten years of progress.

Vreuls, D., & Obermayer, R. W. (1985). Human-system performance measurement in training simulators. *Human Factors*, *27*, 241-250.

Winsch, B. J., Atwood, N. K., Sawyer, A. R., Quinkert, K. A., Heiden, C. K., Smith, P. G., & Schwartz, B. (1993). *Innovative training concepts for use in distributed interactive simulation (DIS) environments.* (Army Project No 627885A 790). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.