Diagnosing Shortfalls in War-Gaming Effectiveness: A Model-Based Approach

Dr. Anna T. Cianciolo
Global Information Systems Technology, Inc.
Champaign, IL
acianciolo@gist-inc.com

William R. Sanders
U.S. Army Research Institute
Fort Knox, KY
william.sanders@knox.army.mil

ABSTRACT

Advances in instructional technology have significantly increased the range of learning environments available to geographically distributed learners. Distributed combat readiness training need not be limited to individual skills now that students can gather in virtual spaces to train on collective skills critical to mission success. However, technological advancement has outpaced the capability to formally assess the quality of collective skills and to diagnose team performance deficits--a capability that is key to transforming learning environments into true instructional systems. In our research, we explored a range of methods for assessing the collective skills of U.S. Army National Guard officers in training as they war-gamed in a virtual tactical operations center. Our intent was to capture aspects of war-gaming performance that would (a) indicate the level of collective skill development; and (b) reveal the causes for team performance shortfalls. We based the design of our assessments on a psychological model of war-gaming that we developed using the results of a comprehensive cognitive task analysis. Our model comprises the psychological constructs associated with effective war-gaming processes and outcomes as well as with key determinants of war-gaming effectiveness. In collaboration with instructors, we administered our assessments to National Guard officers enrolled in the distance-learning version of the Armor Captains' Career Course, taught through the U.S. Army Armor School at Fort Knox, KY. Initial qualitative analysis suggests our war-gaming model and assessments are an appropriate and effective springboard for follow-on research. The primary lesson learned is that training-assessment validation efforts must involve an integrated approach in which assessment design is based on a thorough understanding of the training objectives, assessment implementation is based on a thorough understanding of the learning environment, and assessment validation is accomplished in a controlled setting in which specific hypotheses about the assessments can be tested.

ABOUT THE AUTHORS

Dr. Anna T. Cianciolo has been the Senior Scientist of Instructional Technology at Global Information Systems Technology, Inc. since 2003. She received an M.A. degree in Cognitive & Biological Psychology from the University of Minnesota, and a Ph.D. in Engineering Psychology from the Georgia Institute of Technology. Her research currently focuses on the application of differential psychology and cognitive science to understanding knowledge acquisition and the development of expertise, human performance in teams, and community learning.

William R. Sanders is a Research Psychologist for the U.S. Army Research Institute (ARI) at Fort Knox. He received his M.A. in Industrial and Organizational Psychology from the Ohio State University in 1979. Since joining ARI in 1980 he has served at ARI Headquarters in Washington, U.S. Army Europe, the Army Air Defense School at Fort Bliss, Texas, and the Army Armor School at Fort Knox, Kentucky. His research has supported Air Defense system acquisition programs, and digital system skills acquisition and retention. Current research efforts include development of measurement methods for assessing human performance requirements for future command and control systems.

Diagnosing Shortfalls in War-Gaming Effectiveness: A Model-Based Approach

Dr. Anna T. Cianciolo
Global Information Systems Technology, Inc.
Champaign, IL
acianciolo@gist-inc.com

William R. Sanders
U.S. Army Research Institute
Fort Knox, KY
william.sanders@knox.army.mil

INTRODUCTION

The integration of the virtual learning environment with platforms of collaborative tools has enabled significant advancements in Army distance learning. In today's collaborative virtual learning environments [e.g., the Virtual Tactical Operations Center (VTOC) hosted at the U.S. Army Armor School (USAARS), see Sanders, 2002)], students who are geographically distributed can develop collective skills through synchronous collaboration on practical exercises. However, technological advancement has outpaced the capability to formally assess the quality of collective skills and to diagnose team performance deficits--a capability that is key to transforming learning environments into true instructional systems.

For example, in the Armor Captains' Career Course, Distance Learning version (AC3DL), Army National Guard officers come together in the VTOC for seven practical exercises in which they learn the fundamentals of the military decision-making process and basic company command skills. These practical exercises prepare non-resident students for follow-on resident-phase training at USAARS during which they conduct practical exercises as a co-located group. During both the distributed and co-located phases of collective skills training, AC3DL students each play the role of a battalion-level core staff officer. In these roles, students collaborate on formulating and analyzing combined arms mission plans. AC3DL instructors observe and informally evaluate performance, making corrections in student thinking and debriefing students at the end of each session on how well they performed. The criteria instructors use to evaluate student performance are largely unarticulated, yet the key to formally assessing collective skills and diagnosing performance shortfalls is understanding the basis of their expert judgment.

In the present study, we sought to identify, develop, and validate techniques for assessing the effectiveness of collective skills as demonstrated in a collaborative virtual learning environment. To do this, we worked closely with the students and instructors involved in AC3DL. We focused our efforts on assessing the effectiveness of AC3DL student course-of-action analysis, or war-gaming, so that the implications of our efforts would apply to an area of particular interest and challenge to the training community. Our intent was to capture aspects of war-gaming performance that would (a) indicate the level of collective skill development; and (b) reveal the causes for team performance shortfalls. Although our efforts focused on AC3DL, our goal was not to produce assessments for this particular course of instruction, but to conduct research that would inform the assessment and the development of war-gaming effectiveness more broadly.

A COGNITIVE TASK ANALYSIS OF WAR-GAMING

In order to identify what the war-gaming assessment targets should be, we began our study with a cognitive task analysis of war-gaming. Our task analysis was conducted in three phases: literature review, discussion with subject matter experts (SMEs), and observation of ongoing war-gaming. We sought to determine the cognitive competencies that underlie effective war-gaming and to understand the collective cognitive activities that staff officers engage in during the war-gaming process.

In our literature review, we read several unclassified Army field manuals on mission planning and tactics (e.g., Department of the Army, 2002), articles in the Army professional literature (e.g., Center for Army Lessons Learned, 1998), and a variety of publicly accessible Government-sponsored technical and research reports (e.g., Mullen, Kemper, Harrison, & Bartkoski, 1997), monographs (e.g., Glenn, 1996), and military student theses (e.g., Crain, 1989).

The SMEs we interviewed included current and former observer/controllers at the U.S. Army National Training Center, active and retired lieutenant colonels with battalion command experience, Army schoolhouse instructors, and two retired generals. All of these SMEs had experience as staff officers and as members and/or commanders of a combined arms unit.

Several of these SMEs worked closely with us, providing guidance in assessment design and constructive criticism on draft assessments.

We conducted observations of war-gaming in both a formal instructional setting and during a highly realistic field exercise. The war-gaming instruction we observed was conducted at USAARS with AC3DL resident-phase students and with regular Army officers participating in the resident-version Armor Captains' Career Course. The field exercise was conducted at the National Training Center, during which we observed a brigade staff and an armor task force staff plan stability and security operations. We conducted all of our observations in collaboration with SMEs who provided explanation and insights in real time.

Our task analysis revealed that war-gaming is a mission-planning exercise in which mission events are determined, as opposed to a mission-rehearsal exercise in which mission events are practiced. It is a deliberate, analytical process by which battle staffs refine a commander's course of action by ensuring that critical events on the battlefield are synchronized in time and space and adequately resourced. The desired outcome of this process is a refined plan and a shared understanding among staff officers and their commander of the intended flow of battle and the triggers for the execution of contingency plans.

Our task analysis also revealed that essentially the same general competencies underlie both co-located and distributed war-gaming, although technical skills become more important when communications are digitally mediated. Reference to war-gaming for the remainder of our discussion therefore refers both to co-located and distributed war-gaming.

INDIVIDUAL AND COLLECTIVE WAR-GAMING COMPETENCIES

The individual and collective competencies underlying war-gaming effectiveness are not well understood (though see Olmstead, 1992). That is, the tasks that should be completed for a war game to be considered successful are understood (Mullen et al., 1997), but it is less well known what causes excellent or poor performance of these tasks. Understanding the causes of performance shortfalls is necessary to provide feedback to students and to transform the collaborative virtual learning environment into a true instructional system. We explored the team training literature (e.g., Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995) to gain an understanding of the individual and collective competencies that determine the effectiveness of war-gaming processes and that reflect the execution of effective war-gaming processes. We referred to the same literature to cognitively conceptualize effective war-gaming outcomes.

Determinants of War-Gaming Effectiveness

On the basis of our task analysis and our review of the team training literature, we identified three determinants of war-gaming effectiveness to explore in the present study: critical thinking/analytical reasoning, knowledge of own roles and the roles of others, and tacit knowledge for war-gaming. A brief description of each is presented below.

Critical Thinking/Analytical Reasoning

The purpose of the staff during mission planning and execution is to gather and process information that the unit commander will use as the basis for his decision-making. The critical analysis of incoming information is challenging for staff officers (White, 2001). With the introduction of digital command and control systems, it is now more difficult and more important that they effectively analyze the volumes of incoming data and identify the specific implications of information for mission success (Langley, 2004). The ability to think critically/reason analytically enables staff officers to sift through incoming data with the commander's information requirements in mind, select relevant data for further analysis, and communicate information to the commander in a directed, timely manner. During war-gaming, staff officers must think critically/reason analytically in order to communicate efficiently and effectively with one another and to keep war-gaming activity focused on the commander's intent.

Knowledge of Own Roles and the Roles of Others

During war-gaming, the workflow of the staff is reciprocally interdependent (Tesluk, Mathieu, Zaccaro, & Marks, 1997). That is, the staff officers involved each represent a different functional area (e.g., intelligence, air defense, etc.) and must work closely with one another to produce a refined mission plan that maximally leverages the capabilities in their area of interest. Staff officers contribute to refining the plan by reporting the implications of the current situation in their area of interest for how mission events will play out. Knowledge of one's own role and the roles of others on the staff enables staff officers to understand each other's information needs and to effectively meet those needs (Cannon-Bowers et al., 1995; Olmstead, 1992).

Tacit Knowledge for War-Gaming

The procedure for conducting a war game and its desired outcome is explicitly described in doctrine (DOA, 2002). However, conducting an effective war-game requires an understanding of the process that goes beyond knowing explicit procedure. Staff officers must understand the purpose of war-gaming in order to engage in many of the tasks that are required for an effective war game. Specifically, they must understand that producing doctrinal war-gaming products, such as the synchronization matrix, is not the goal of war-gaming but a record of its results and that the quality of the results is dependent upon an integrated staff effort.

Tacit, or experience-based knowledge for war-gaming provides staff officers with a more nuanced understanding of the war-gaming process than is explicitly stated in doctrine, and enables them to translate this understanding into action through analysis tasks during the war game. For example, a less experienced staff may estimate combat losses without further analysis because the estimate is enough to fill out a cell in the synchronization matrix. A more experienced staff would analyze what implications combat losses have for the commander's decision making.

Effective War-Gaming Processes

Effective war-gaming processes are enabled when staff officers possess critical/analytical skills, recognize each other's information needs, and understand the purpose of war-gaming. Staff officers prepared in this manner will communicate better with one another and think more adaptively during the war-gaming process. We describe these war-gaming processes in more detail below.

Team Communication

Because war-gaming involves a reciprocally interdependent workflow (Tesluk et al., 1997), effective information sharing among staff officers is critical for producing a refined course of action. For each task that must be completed during a war game, staff officers must share specific information about their area of interest such that all functional areas are integrated into the refined plan. Effective information sharing eludes many staffs, with some staff officers routinely left out of the war-gaming process.

Adaptivity of Team Thought

Automatized adaptive-thinking skills have been recognized as a critical aspect of expert command decision-making (Lussier, Shadrick, & Prevou, 2003). Expert commanders are tuned into their environment and recognize the implications of particular environmental conditions for making rapid decisions in response to unforeseen events. Staff officers must anticipate unforeseen events during mission execution by evaluating during war-gaming the same factors that the commander must consider when he makes decisions. In this way, war-gaming supports the commander's decision making by ensuring that the manpower and resources can be available when decisions must be made. Adaptivity of team thought among staff officers enables them to consider "What if?" on the basis of conditions present in the battle situation and to make contingency plans to address possible events during execution. This creates a flexible plan that is responsive to a range of mission events.

Effective War-Gaming Outcomes

As stated previously, the desired outcome of war-gaming is a refined plan and a shared visualization of the intended flow of battle and the triggers for the execution of contingency plans. This outcome stems from effective communication and adaptive thinking among the staff officers. Evaluation of the quality of a particular war-gaming outcome therefore involves assessing shared battle visualization and the degree to which the mission plan has integrated the capabilities of each staff officer's functional area.

Shared Battlefield Visualization

Shared battlefield visualization can be characterized as shared situation awareness (SA, e.g., Endsley & Smolensky, 1988). It is (a) awareness of the mission plan elements and their locations (Level 1 SA); (b) understanding of how plan elements are synchronized in time and space and the implications of their success/failure for one another (Level 2 SA); and (c) projection of how these elements will function as the plan is executed and where key decisions will need to be made (Level 3 SA). Shared SA is a critical outcome of war-gaming because it is the means by which the command and control team anticipate each other's actions during mission execution.

Integrated Mission Plan

The integrated mission plan maximally leverages the capabilities of each staff officer's functional area by synchronizing the efforts of each area in time and space. The integrated mission plan is a critical outcome of war-gaming because its development requires careful consideration of the status and constraints of each functional area and the implications of these factors for how mission events can play out. This consideration enhances shared SA and reduces the risk of fratricide resulting from uncoordinated mission events.

ASSESSMENT OF WAR-GAMING COMPETENCE

In this section, we describe the assessments that we designed to capture all three aspects of war-gaming performance (determinants, processes, outcomes). Where applicable, psychometric properties of the assessments are discussed.

Critical Thinking/Analytical Reasoning

To assess staff officers' ability to identify and communicate relevant information, we developed the Mission Analysis Briefing Exercise (MABE). The MABE is a computer-based adaptation of a classroom introductory exercise developed by an instructor at the U.S. Army Command and General Staff College. In this exercise, students are asked to review a set of PowerPoint slides comprising a hypothetical mission analysis briefing. Students are told that the briefing is too long and must be shortened from 36 to 15 slides in order to communicate the most important information in the briefing. They are given 15 minutes to complete the exercise. The score for this exercise is the number of slides in the abbreviated briefing containing relevant information divided by the total (reduced) number of slides (possible range = 0-1). Slides containing relevant information were determined in collaboration with the instructor who designed the exercise.

Knowledge of Own Roles and the Roles of Others

To assess student understanding of the information needs of core battalion staff officers (the officers whose roles they would play during war-gaming), we designed a 42-item assessment in which each item presented a key war-gaming task (see Mullen et al., 1997). Students were asked to indicate which of nine staff officers must share information in order to perform the task effectively. For each item, the score is the number of staff officers correctly identified divided by the number of staff officers who should have been identified (possible range = 0-1). The total score is the average of the scores for each of the 42 items (possible range = 0-1). Answers to each item were determined by review of doctrine and SME input.

In initial pilot testing, the split-half reliability of the Staff Roles Assessment was quite high (r_{xx} = .98), so we reduced it by half in order to ease the workload on the examinee. The internal-consistency reliability of the reduced assessment is also quite high, at .95.

Tacit Knowledge for War-Gaming

We designed a short, five-question multiple-choice quiz to capture the level of development of students' understanding of the purpose of war-gaming. Each question asks about the purpose of some aspect of war-gaming. For example, one question asks: "The war-gaming process is conducted using multiple iterations of an action-reaction-counteraction (ARC) cycle. Why?" Correct answers were determined by SME input. The score for this assessment is the number correct minus .25 times the number incorrect.

The internal consistency reliability for the War-Gaming Purpose Quiz is .38 (*N* = 20). This estimate is quite low given the quiz is intended to assess a single construct. However, the quiz is only five questions long and the length of an assessment has implications for its reliability. When the Spearman-Brown formula is used to determine what the reliability of the War-Gaming Purpose Quiz would be if it was as long as the abbreviated Staff Roles Assessment (21 items), we derive a reliability of .72.

Team Communication

To assess information sharing, we selected a subset of 12 tasks from the Staff Roles Assessment that we knew could be addressed during AC3DL war-gaming. We used these twelve tasks to design an observer checklist that was based roughly on the TARGETS methodology (Fowlkes, Lane, Salas, Franz, & Oser, 1994). We used doctrine and SME review to identify the information that must be shared in order for each of the 12 tasks to be completed effectively. An example war-gaming task and its associated information-sharing requirements is shown below in Table 1.

Table 1. Example Item from the Team Communication Checklist

Task	Information Shared	Rating
Determine High Priority Targets	Templated location of the enemy and key enemy assets (e.g., artillery, C2 nodes)	
	Concept of maneuver	
	Resupply rates for select munitions	

A key difference between our checklist and a TARGETS checklist is that the occurrence of tasks to be observed is dependent on the team's effectiveness, rather than on the design of the team-performance situation (i.e., more effective teams would attempt more tasks). This design allows us to capture the completion of war-gaming tasks independently from the effectiveness of team communication, but provides a means for focusing observer attention. A second difference is that our checklist requires a greater degree of observer judgment because ratings of information sharing were not binary (i.e., information shared/not shared). Raters use 0, .5 or 1 to indicate the degree to which information was shared. Similarly, each task is rated 0, .5 or 1 for its level of completion.

The task score is the average task rating for all 12 tasks (possible range = 0-1). The information-sharing score is the sum of the average item ratings within each of the twelve tasks, divided by 12 (possible range = 0-1). The total score is sum of the task score and the information-sharing score divided by two (possible range – 0-1), representing a combined assessment of the accomplishment of war-gaming tasks and team communication.

Where two raters used the Team Communication checklist, they agreed on 67% of the judgments that a war-gaming task had been attempted. The correlation of the scores they assigned to each task was .54. Similarly, these same two raters agreed on 77% of the judgments that an attempt to share a particular piece of information occurred. The correlation between the scores they assigned to information sharing was .69. Differences among the two raters appear to come from differing levels of experience observing AC3DL war-gaming and differing levels of familiarity with the doctrinal information requirements for each war-gaming task.

Adaptivity of Team Thought

To assess students' adaptive-thinking skills during the war game, we created an interactive observer checklist based on the Think Like a Commander (TLAC) training method (Lussier et al., 2003). Using the mission scenario war-gamed in the AC3DL course, we identified "What if?" questions about the mission that observers could ask students during the war game. As students answer a question posed by the observer, the observer rates the quality of student responding on a scale of 0-2. Anchors for the rating scale were determined by SME input, with example "0-" and "2-quality" answers for each question provided to the observer. An example question and its rating scale anchors are shown below in Table 2.

Table 2. Example TLAC Checklist Item

Question	Example 0	Example 2
What cover and concealment can the enemy take advantage of in order to deceive 1-22 CAV regarding the size/strength of threat forces?	Solutions that do not include consideration of local populace	Use the civilian population to blend into the terrain
	Solutions that are not based on fighting an asymmetric enemy (e.g., students think in terms of large units, heavy equipment, etc.)	Disperse widely, remain in small teams, and move frequently such that effective estimates require an integrated intel effort

Two "What if?" questions correspond to one of the eight TLAC expert-thinking themes (e.g., model a thinking enemy, consider the effects of terrain), totaling 16 questions in the checklist. The score for the TLAC Checklist is the average of the question scores earned by the students (possible range = 0-2).

Shared Battlefield Visualization

To assess shared SA as a war-gaming outcome, we devised an exercise based on an Army doctrinal war-gaming product called a decision support matrix (DSM). The DSM requires staffs to consider where on the battlefield the commander will have to make a decision during mission execution, what information he needs to make those decisions, and the options available to the commander at the time of decision making. As shown in Table 3, our SA exercise requires examinees to answer questions associated with elements of the DSM, which correspond to each level of SA. There were five sets of these questions in the exercise.

Table 3. Example SA Exercise Questions

SA Level	Question
1	What is the center of mass for NAI3?
2	If the enemy strongpoint is located forward on high ground at NAI3, what would this reveal about the enemy intends to do?
3	What should TF 1-93 do if the eastern enemy strongpoint is located forward on high ground?

We used the course of action that the AC3DL students war-game in class as the basis for our SA exercise questions. With the help of subject matter experts, we developed an exercise "key" we could use to score student war-gaming outcomes. The individual score for this exercise is the sum of correct answers, weighted according to the level of situation awareness required to answer the question correctly (possible range = 0-30). The team score for this exercise is the percent agreement on *correct* answers for each question weighted according to the level of SA and summed (possible range = 0-30).

Integrated Mission Plan

To assess the level of integration of the mission plan, we designed an Integrated Overlay Exercise. In this exercise students are asked after they complete the war game to create a graphical overlay of each phase of the refined mission plan. The overlay is to contain the key elements of friendly activity not represented in the course of action sketch (i.e., the activity determined during war gaming), including logistical assets and control measures. Upon completing the overlay, students are asked to brief the overlay to an observer, providing a rationale for the placement of each element. Students earn one point for each element they include in the overlay (max = 21), compared to a "key" created by a SME. The observer rates student rationale for each element using a checklist in which the components of rationale for each element are listed. An example element and its rationale components are shown in Table 4.

Table 4. Example Integrated Overlay Checklist Item and Rationale

Element	Rationale Component
Smoke	Smoke at this location will mask TF 1-93 maneuver and protect movement to attack positions
	Smoke location is coordinated with S2 projections for weather/wind direction

The score for each element is the proportion of the total number of rationale components provided by the student (possible range = 0-1). The score for the exercise is the number of elements included in the overlay divided by the sum of the element rationale scores (possible range = 0-1).

ASSESSMENT VALIDATION

Validation of training-performance assessments must demonstrate that the assessments reliably capture the training objectives they are intended to capture. That is, the assessments must (a) differentiate between more and less capable students as identified by independent criteria; and (b) reflect the effect of learning as identified by independent criteria.

We administered subsets of the assessments we developed to five independent groups of AC3DL students during one or both occasions they war-gamed (i.e., in a virtual learning environment and/or in the classroom). An overview of the assessment administration is shown in Table 5. Additionally, we administered a demographic survey to all groups and asked instructors to rate student performance after each war game (relative to doctrinal standards and relative to other student groups) in order to capture external criteria reflecting more and less experienced or capable student groups.

Table 5. Assessment Administration

Group	Instructor	Learning Envt.	Assessments Administered
1	A	Virtual	N/A
1	A	Classroom	Staff Roles, War-Gaming Purpose, Team Communication
2	B	Virtual	MABE, Staff Roles, War-Gaming Purpose, Team Communication
2	A	Classroom	TLAC Checklist
3	A	Classroom	Staff Roles, TLAC Checklist
4	A	Virtual	MABE, Staff Roles, War-Gaming Purpose, Team Communication, SA Exercise
5	B	Virtual	MABE, Staff Roles, War-Gaming Purpose, Team Communication, SA Exercise

Review of Table 5 reveals that the administration of our assessments in the present study does not lend itself well to statistically testing hypotheses about the validity of our assessments. That is, due to time constraints, no student group could be given the same assessment twice. Due to technical and other constraints some assessments could only be administered to some groups or not at all. Finally, two different instructors provided independent performance evaluations for various student groups and a different scenario was used for war-gaming in each of the different learning environments. For these reasons beyond our control it was not possible to empirically explore the validity of the assessments beyond basic psychometric properties.

That said, the present effort represents an extension of best practice in the understanding and assessment of war-gaming effectiveness, and serves as a springboard for follow-on research. The competencies underlying war-gaming effectiveness have not heretofore been identified or systematically represented. Moreover, the psychometric properties of assessments for which sufficient data were available (Staff Roles Assessment, War-Gaming Purpose Quiz, and Team Communication Checklist – described above), were satisfactory. In addition, performance on these assessments reflected expected patterns of responding given what is known about war-gaming performance in field and operational settings.

First, without exception, students who took the Staff Roles Assessment most accurately identified the information needs of the operations officer and the intelligence officer, who are a central focus of the war-gaming process (mean percent correct for these officers was, respectively, 29 and 9 percentage points above the mean percent correct for the entire assessment). The information needs of other staff officers, i.e., those representing functional areas that deal with combat service support, were poorly understood by the students we worked with just as they are not well understood by many combat arms officers in general (CALL, 1998; mean percent correct for these officers was, on average, 15 percentage points below the mean percent correct for the entire assessment).

Second, students generally earned low scores on the assessments. These low scores do not reflect a poorly functioning educational system in the U.S. Army, but rather a lack of student experience combined with the challenge of war-gaming in a setting that is not naturalistic. AC3DL students are junior officers in the National Guard and have generally had no prior experience with battalion-level mission planning. In addition, the schoolhouse environment necessitates stripping important context away from the typical war-gaming environment. Indeed, low scores on these assessments were achieved as expected. On a positive note, the assessments represent essential war-gaming competencies, and as such support the development of training to address the noted deficiencies in experience.

LESSONS LEARNED

The primary lesson learned from this study is that training-assessment validation efforts must involve an integrated approach in which assessment design is based on a thorough understanding of the training objectives, assessment implementation is based on a thorough understanding of the learning environment, and assessment validation is accomplished in a controlled setting in which specific hypotheses about the assessments can be tested. The integrated approach enables training researchers to empirically validate performance assessments while at the same time evaluating the feasibility of the assessments and ensuring their relevance in the actual training environment.

In our study, we did not have the liberty to work with a large number of research participants in a controlled setting, as is recommended for assessment-validation studies. Moreover, because we worked with a small number of student groups from just one course (AC3DL), our ability to field-test certain hypotheses about the validity of the assessments was significantly limited. However, we were able to determine what the nature of assessments for capturing war-gaming performance should be (i.e., what competencies and behaviors should be assessed) and to evaluate the feasibility of implementing these assessments in a current collaborative virtual learning environment.

Important technical factors to consider when implementing performance assessments in collaborative virtual learning environments include the speed of the Internet connection used by students in the course, the operating system and service packs in use on student computers, and the stability of the learning environment. The reality of advanced training technology is that it is not similarly experienced by all students and that difficulty with various elements of the technology--due to unforeseen and non-replicable conditions--is commonplace. These factors influence the feasibility of all types of assessment, including observer checklists and automated data collection. To ensure that assessments capture what they are intended to capture, designers and developers must make certain that the assessments will be accessible to the range of computer configurations available to students and robust in the face of technical difficulty.

Important behavioral factors to consider when implementing performance assessments in collaborative virtual learning environments include the time required to administer/score the assessments, the usability of the assessments, the level of experience of the students taking the assessments, and the level of buy-in with which both students and instructors approach the assessment methods and results. These factors influence whether the assessments are actually administered and/or whether the results of the assessments are shared and discussed. To ensure that assessments will be used, designers and developers must work closely with instructors to make certain the assessments (a) capture constructs or behaviors of interest to the instructors; (b) feature an appropriate level of difficulty, given student experience levels; (c) fit well with the existing curriculum and tone of the course; (d) require minimal extra learning to administer/take; and (e) produce scores and feedback that are meaningful to instructors and students.

CONCLUSIONS

The promise of collaborative virtual learning environments is that they will support the development of collective skills among students who are not co-located. This technical capability may ultimately support combined arms training in which students located at Army schoolhouses nationwide conduct collaborative exercises in one virtual space. The development of assessment methods to capture collective performance and to diagnose shortfalls ensures that the learning environment serves as a true training system. The present work represents an initial attempt to identify the determinants and indicators of war-gaming performance, a critical collective activity during Army mission planning. We have identified key competencies that should be assessed, implemented assessments to capture these competencies, and conducted a preliminary validation of these assessments. Continuation of this effort should focus on controlled validation studies of assessments of war-gaming competence to identify best practice in design and implementation.

ACKNOWLEDGEMENTS

This project was funded by the U.S. Army Research Institute for the Behavioral and Social Sciences. We gratefully acknowledge LTC (Ret.) Jim Harrison for his generous assistance, mentorship, and patience throughout this project. We also gratefully acknowledge LTC (Ret.) Glen DeWillie, BG (Ret.) Bill Mullen, III and LTC (Ret.) George Hodge for their assistance in the development and review of our assessments.

REFERENCES

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R. A. Guzzo, E. Salas, and Associates (Eds.), *Team effectiveness and decision making in organizations* (pp. 333-380). San Francisco, CA: Jossey-Bass.

Center for Army Lessons Learned (CALL). (1998). *NTC trends compendium: 1QFY98 through 2QFY98*. Fort Leavenworth, KS.

Crain, W. F. (1989). *Battle staff operations: Synchronization and planning at battalion and brigade level*. Unpublished Master's thesis. U.S. Army Command and General Staff College, Fort Leavenworth, KS.

Endsley, M. R., & Smolensky, M. W. (1998). Situation awareness in air traffic control: The picture. In M. W. Smolensky & E. S. Stein (Eds.), *Human factors in air traffic control* (pp. 115-154). San Diego, CA: Academic Press.

Fowlkes, J. E., Lane, N. E., Salas, E., Franz, T., & Oser, R. (1994). Improving the measurement of team performance: The TARGETs methodology. *Military Psychology, 6*(1), 47-61.

Glenn, H. C., III. (1996). *Building the tactical nerve center: Enhancing battalion commander and staff performance in the tactical decision making process*. Monograph of the School of Advanced Military Studies, U.S. Army Command and General Staff College, Fort Leavenworth, KS.Langley, J. A. G. (2004). Network-centric warfare: An exchange officer's perspective. *Military Review*, November-December, 47-52.

Lussier, J. W., Shadrick, S., & Prevou, M. I. (2003). *Think like a commander prototype: Instructor's guide to adaptive thinking* (Research Product No. 2003-02). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Mullen, W. J., III, Kemper, T. R., Harrison, K. E., & Bartkoski, T. P. (1997). *Plan for combat operations (Battlefield Function 18) as accomplished by a heavy brigade* (Vol. 2, Assessment Package, Research Product No. 98-04). Alexandria, VA: U.S. Army Research Institute for the Behavioral & Social Sciences.

Olmstead, J. A. (1992). *Battle staff integration* (IDA Paper P-2560). Alexandria, VA: Institute for Defense Analysis.

Sanders, W. R. (2002). *Collective staff training in a virtual learning environment* (Research Report No. 1788). Washington, DC: U.S. Army Research Institute of the Behavioral & Social Sciences.

Tesluk, P., Mathieu, J. E., Zaccaro, S. J., & Marks, M. (1997). Task and aggregation issues in the analysis and assessment of team performance. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, research, and applications* (pp. 197-224). Mahwah, NJ: Lawrence Erlbaum Associates.

U.S. Department of the Army. (2002, July). *FM 5-0: Army planning and orders production* (Final Draft). Washington, DC.

White, D. (2001, May). The air defense commander and CCIR: Rather than critical information requirements analysis to anchor their decisions, commanders are treated to a fire hose of data. *ADA Magazine*.