Leveraging Commercial Video Game Technology to Improve Military Decision Skills

Holly C. Baxter, Ph.D., Karol G. Ross, Ph.D., Jennifer Phillips, Jennifer Shafer
Klein Associates Inc.
Fairborn, Ohio
holly@decisionmaking.com
karol@decisionmaking.com
jshafer@decisionmaking.com

Jennifer Fowlkes
Chi Systems
Orlando, Florida
jfowlkes@chisystems.com

ABSTRACT

Desktop simulations and digital game-based technologies have earned much attention for their potential as training interventions. Supporters view the physical realism and interactivity of the technologies as a powerful means of fostering the development of cognitive skills. The USMC is leveraging aspects of digital-game based methods and inserting them into tactical decision-making simulations (TDSs) that can supplement existing training. This paper presents a pilot study conducted to examine the training utility of TDSs. One TDS, Close Combat Marine, was assessed against traditional paper-based Tactical Decision Games (TDGs). The objectives of the study were to examine the utility of metrics for assessing improvement in cognitive skills as a result of TDSs and TDGs, and also to generate initial hypotheses regarding optimal ways of designing and implementing TDSs. Eight separate metrics were employed, including final exam scores, knowledge tests, surveys, cognitive assessments, and behaviorally anchored rating scales that assess mental models for tactical thinking. Results indicate that students in the TDS intervention were more motivated to engage in the training sessions than those in the TDG group. The findings are consistent with prior research indicating higher levels of learner motivation associated with digital game-based technologies. The results also indicate that TDSs and TDGs may be differentially beneficial. In this pilot study, for example, the TDG intervention seemed better at training mental simulation, planning, and mission focus. The TDS intervention seemed to be stronger at addressing timing considerations, execution knowledge, and team coordination. Further research examining a broader range of TDSs should focus on more accurately gauging their relative advantages and disadvantages and on guiding their incorporation into existing training programs. This research was funded by the Office of Naval Research under contract #N61339-99-D-0012 to CHI Systems, Inc. under the USMC Program Manager, Training Systems Science and Technology Division..

ABOUT THE AUTHORS

Holly C. Baxter, **Ph.D.** is a Research Associate II specializing in Instructional Design, Evaluation Metrics, and Training. As a member of Klein Associates, she has worked in each of these areas on a variety of projects including the areas of knowledge management, training simulation technologies, enhancing situation awareness, and developing metrics for understanding the effectiveness of simulation technologies. Dr. Baxter holds a BA in Communication from the University of Dayton, a MA in Organizational Communication and Training from Indiana University, and a Ph.D. from Indiana University in Organizational Communication and Management with a focus on Training & Instructional Design.

Karol G. Ross, Ph.D. is a Senior Research Associate at Klein Associates Inc., currently pursuing research in decision skills training using scenario-based training. As a Research Psychologist for the U.S. Army Research Laboratory, she conducted research and development for adaptive thinking training at the Brigade level. She also served as a Senior Researcher for BDM International studying battle staff performance and as a Research Psychologist for the Army Research Institute working in the area of large-scale simulation training for echelons above Corps. She earned her doctoral degree from the University of Tennessee in Experimental Psychology in 1984.

Jennifer Phillips, a former Research Associate at Klein Associates Inc., conducts research into the nature and acquisition of expertise and decision-centered training. Research domains have included training for small unit leaders in military operations in urban terrain, Air Force personnel for Operations Other Than War, and sensemaking skills in Information Operations. Ms. Phillips received a B.A. in Psychology from Kenyon College in 1995.

Jennifer Shafer is a Research Analyst II at Klein Associates. Ms. Shafer is currently co-leading a project assessing different knowledge representation techniques. She is also supporting a project funded by the AFRL to test the robustness of a model that describes cultural variance in cognition (the Cultural Lens Model) and is coordinating Klein Associates' workshops in Cognitive Task Analysis. Ms. Shafer holds an M.A. in Neuroscience from Johns Hopkins University, Baltimore, MD, and B.S. degrees in Psychology and Biology from the University of Washington, Seattle, WA.

Jennifer E. Fowlkes, Ph.D. is Senior Cognitive Engineer with CHI Systems, Inc. She has over fifteen 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 has focused on measuring adaptive team performance in distributed training environments. Dr. Fowlkes holds a Ph.D. in Experimental Psychology from the University of Georgia.

Leveraging Commercial Video Game Technology to Improve Military Decision Skills

Holly C. Baxter, Ph.D., Karol G. Ross, Ph.D., Jennifer Phillips, Jennifer Shafer
Klein Associates Inc.
Fairborn, Ohio
holly@decisionmaking.com
karol@decisionmaking.com
jshafer@decisionmaking.com

Jennifer Fowlkes
Chi Systems
Orlando, Florida
jfowlkes@chisystems.com

Desktop simulations and digital game-based technologies have earned much attention for their potential as training interventions. Supporters view the physical realism and interactivity of the technologies as a powerful means of fostering the development of cognitive skills. The U.S. Marine Corps (USMC) is leveraging aspects of digital-game based methods and inserting them into tactical decision-making simulations. Is the promise of game-based training being realized in the arena of complex cognitive skills? This paper presents a pilot study conducted to examine the training utility of a specific TDS, Close Combat Marine (CCM) in a leadership course, as compared to traditional paper-based Tactical Decision Games (TDGs) already in use in the course. The objectives of the study were to examine the utility of metrics for assessing improvement in cognitive skills as a result of TDSs and TDGs and to generate initial hypotheses regarding optimal ways of designing and implementing TDSs.

The training evaluation was implemented within the eight-week USMC Infantry Platoon Sergeant (IPS) Course taught at the Advanced Infantry Training School at Camp Geiger, North Carolina. The intent was to compare the relative impact of these forms of decision-making training on measures of attitudes toward the training and on changes in tactical decision-making skills and knowledge. Two components of the evaluation were developed to facilitate the identification of training strengths and weaknesses: a taxonomy of competencies and a multi-measure, multi-level evaluation approach.

TAXONOMY OF COMPETENCIES

We first identified competencies that we believed should be addressed by a tactical skills training program. Our resulting taxonomy of cognitive skills provided a focus for the development of evaluation methods. While the use of a taxonomy provides the framework for a diagnostic assessment of a program's strengths and weaknesses, it can also serve as an organizing framework as data on effectiveness of TDSs are accumulated across studies. Two classifications of the competencies were used: "macrocognition" functions and processes drawn from Klein Associates' research and dimensions of tactical thinking that have been used in a number of studies of tactical thinking training.

Macrocognition

Macrocognition is a term used to describe the array of cognitive activities performed in naturalistic settings. It describes the emergent cognitive functions and processes that arise in naturalistic settings (Klein, et al., 2003). Klein et al. (2003) identified a core set of macrocognitive functions: naturalistic decision making, sensemaking/ situation assessment, planning, adaptation/replanning, problem detection, and coordination. In service to these core functions is a set of macrocognitive processes: maintaining common ground, developing mental models, uncertainty management, turning leverage points into courses of action, attention management, and mental simulation and story building. For any particular cognitively complex performance, a subset of the macrocognitive components will be most prevalent. The components that seemed to be most relevant, and those targeted for evaluation in this training, were coordination, mental simulation, uncertainty management, planning, decision making, and leverage points. (See Table 1.) These cognitive functions and processes provided the basis for paper-based assessments and observations.

Dimensions of Tactical Thinking

Researchers have identified eight dimensions or themes of tactical thinking (Ross, & Lussier, 1999; Lussier, Ross, & Mayes, 2000) based on initial cognitive task analysis of expert tacticians (Deckert, Entin, Entin, MacMillan, & Serfaty, 1994).

Table 1. Definitions of Macrocognitive Components

Macrocognitive Activity	Definition
Coordination	Coordination is the attempt by multiple entities to act in concert. Its purpose is to achieve a common goal by carrying out a shared script or plan.
Mental Simulation	Mental simulation is the process for consciously enacting a sequence of events, such as imagining how a Course Of Action (COA) will play out in the future.
Uncertainty Management	Uncertainty is what we do not know or understand about a given situation, defined as "doubt that threatens to block action." Uncertainty involves situations in which key information is missing, unreliable, ambiguous, inconsistent, or too complex to interpret, resulting in a reluctance to act.
Planning	Planning is the process of contemplating and devising actions for some future execution following a decision.
Decision Making	Decision making is the identification of a feasible COA from experience accumulated in similar situations; it may involve, but does not require, a comparison of the strengths and weaknesses of alternative COAs.
Leverage Points	Leverage points are opportunities for making critical changes at a relatively low effort, and a means by which COAs are generated.

The eight themes and a brief description are as follows (Shadrick & Lussier, 2002):

- *Keep a Focus on the Mission and Higher's Intent*--Commanders must never lose sight of the purpose and results they are directed to achieve, even when unusual and critical events may draw them in a different direction.
- *Model a Thinking Enemy*--Commanders must not forget that the adversaries are reasoning human beings intent on defeating them. It is tempting to simplify the battlefield by treating the enemy as static or simply reactive.
- *Consider Effects of Terrain*--Commanders must not lose sight of the operational effects of the terrain on which they must fight. Every combination of terrain and weather has a significant effect on what can and should be done to accomplish the mission.
- *Use All Assets Available*--Commanders must not lose sight of the synergistic effects of fighting their command as a combined arms team. They consider not only assets under their command, but also those which higher headquarters might bring to bear to assist them.
- *Consider Timing*--Commanders must not lose sight of the time they have available to get things done. Experts have a good sense of how much time it takes to accomplish various battlefield tasks. The proper use of that sense is a vital combat multiplier.
- *See the Big Picture*--Commanders must remain aware of what is happening around them, how it might affect their operations, and how they can affect others' operations. A narrow focus on your own fight can get you or your higher headquarters blind-sided.
- *Visualize the Battlefield*--Commanders must be able to visualize a fluid and dynamic battlefield with some accuracy and use the visualization to their advantage. A commander who develops this difficult skill can reason proactively like no other. "Seeing the battlefield" allows the commander to anticipate and adapt quickly to changing situations.
- *Consider Contingencies and Remain Flexible*--Commanders must never lose sight of the old maxim that "no plan survives the first shot." Flexible plans and well thought out contingencies result in rapid, effective responses under fire.

LEVELS OF EVALUATION

The second component of the evaluation was to collect measures at multiple levels. The evaluation levels we used are based in part on a hierarchy presented by Kirkpatrick (1976) and expanded by Kraiger, Ford, and Salas (1993). Level 1, pre-course assessment, serves primarily as a baseline for post-training assessments. This level includes demographic information and pre-course expectations. Level 2 assessments focus on trainee opinions of the training and attitudes towards principles taught in the training. Positive trainee reactions to training and attitude change in the desired direction are taken as evidence of an increased awareness of principles being trained and a likely increase in willingness to use principles appropriately on the job.

Level 3 focuses on learning--evidence that trainees have learned concepts related to tactical decision making. Level 4 is the assessment of skills within context-rich settings such as field exercises, simulations, or on the job. To the extent that principles are "internalized" by trainees, there should be differences between trained and untrained teams or individuals in their behavior in situations that require application of these principles. Level 5 provides assessment of organizational benefits such as accident reduction and improved productivity. The present evaluation primarily focused on evaluation levels 1 through 3, although some performance (level 4) was assessed.

METHOD

Participants

Participants were 14 students attending the IPS course. Because of the small number of participants, students in the TDG and TDS groups were matched on background (years in service) and experience (military occupational specialty) variables in order to mitigate potential group differences at the outset of data collection.

Materials and Procedures

Close Combat Marine (CCM)

Close Combat Marine is a computer-based combat simulation developed by Atomic Games, Inc. The focus of the simulation is on infantry combat at the small-unit level. The 2-CD set provides the game software as well as training documentation and guidance. In addition, the USMC training and collective and individual standards that should be incorporated into training are included with the CD.

Tactical Decision Game

The TDGs were based on the same scenarios as those of CCM. Both were offensive tactics scenarios and both utilized the same map. To implement the TDGs, instructors presented the tactical scenario and map. Students developed a COA within a 15-minute time period and then briefed it to the instructor and student group using a sand table. The instructor and students critiqued the student's plan and provided additional "what-ifs" to consider. The TDGs required approximately 1.5 hours to complete.

Demographic Questionnaire

Participants completed a demographic questionnaire to collect information on military background and computer game usage.

Training Expectations Questionnaire

When a trainee's expectations of a training program are not met, the effectiveness of the training may be lessened. Trainee expectations were collected before and after the training intervention.

Scenario-Based Cognitive Assessment

The scenario-based cognitive assessment tool assessed improvement in tactically relevant mental models as a result of the TDS and TDG interventions. The scenario-based cognitive assessment tool consists of a pre-test and a post-test package. Each package is anchored around a TDG scenario that consists of a sketch map and a narrative description of the general situation.

In each testing period, the students were required to read the scenario material and then answer five questions, putting themselves in the role of the unit commander. The instrument could be administered in about 45 minutes. The questions were:

1. What actions would you initiate immediately and what information would you seek? [5 minutes]
2. State the frag order for your plan (including your intent) and give a brief rationale. [10 minutes]
3. What are the risks/vulnerabilities to your plan and how would you counter them? [5 minutes]
4. Did you consider other courses of action? Why did you decide not to use them? [5 minutes]
5. What are the possible enemy courses of action and his intent? Why do you think that is his intent? [5 minutes]

The assessment tool used to rate the performance is analogous to a behaviorally anchored rating scale (BARS) (Muchinsky, 2003; Riggio, 2000). A tactical thinking BARS framework for evaluating each question was given to a subject-matter expert (SME) who was blind to the students' experimental condition. The framework describes each of four themes that were identified prior to testing as relevant to both the pre- and post-scenarios: Model a Thinking Enemy; Consider Effects of Terrain; Know and Use All Available Assets; and Consider Timing. Each theme for each student was rated on a five-point scale for which descriptors or anchors had previously been identified.

Course of Action (COA) Exercise
The COA exercise required each student to devise a COA for the company commander's order that provided the basis of the TDG/TDS session. The exercise, which was paper-based, was administered immediately prior to (pre-test) and immediately following (post-test) each TDG/TDS session. The pre-test required student response to three questions; the post-test required a response to five. This measure enabled the assessment of changes to each student's COA as a result of the TDG or TDS interventions.

A USMC SME, blind to the experimental condition, scored each pre- and post-test response using the COA Exercise scoring form, which consisted of four questions designed to measure the quality of each student's COA. The scoring form also contained eight questions designed to capture the nature of the changes to each student's post-test response compared to his/her pre-test response.

Knowledge Tests
Three knowledge tests were used to assess student declarative knowledge and the application of that knowledge.

- *Post Exercise Knowledge Test.* After every TDS/TDG, students completed a three- to six- item knowledge test asking them to apply knowledge obtained during lecture to the training scenario. The test items were developed to sample training objectives from the Offensive Tactics, Defensive Tactics and Patrolling lectures.
- *Multiple Choice Knowledge Test*. At the end of the IPS course, students took a 100-item final exam. Final exam scores were obtained and used to support the evaluation.
- *Combat Orders Test*. As part of their training in the IPS course, students developed combat orders for each of the targeted lectures. Rating scales were developed to obtain instructor ratings in five areas from the combat orders.

Instructor and Student Surveys
The course instructors and students completed an instructor or student survey at the end of the TDG/TDS portion of the IPS course. The purpose of the survey was to collect instructor and student opinions regarding how well and in what ways the TDG and TDS sessions supported course-learning objectives.

Classroom Observation
Researchers recorded classroom interactions and discussions for each TDG and TDS session.

- *Tactical Thinking*. Using the BARS tool, observers coded the types of discussions that occurred in each of the sessions. The objective was to assess which dimensions of tactical thinking predominate under the different training formats and to determine whether there is variation in the level of the discussions across the different formats.
- *Classroom Interactions*. Observers counted the number of student-to-student and instructor-to-student interactions during the TDG and CCM sessions. The objective was to determine whether one format encouraged more interaction and collaboration than the other.

RESULTS

Training Expectations Questionnaire

A mixed model analysis of variance was performed on each of the expectations questionnaire items, with administration time as the within-subjects factor (pre versus post) and group (TDG versus TDS) as the between-subjects factor. Two significant differences were found. At the outset of the training, the TDS group was more likely to disagree with the statement, "I don't think I will ever use the information I learn here." Thus, they may have had higher expectations for the usefulness of the training compared to the TDG group. This difference was maintained on the post-training measurement. The other difference was that both groups declined in their agreement with the statement, "I exerted considerable effort to improve my skills in the training." This suggests, perhaps, less motivation at the end of training was found in both groups.

Student Survey

The student survey responses were analyzed utilizing two-tailed independent samples t-tests performed on each of the questionnaire items. Table 2 shows results for items that reached statistical significance ($p < .05$) and for trends ($p < .1$).

Trainees in the TDS group found the sessions to be significantly less boring than the TDG participants. In addition, those in the TDS group rated themselves as significantly more likely to participate in future sessions than those in the TDG group. Trainees in the TDS group were more likely to indicate that training improved their ability to coordinate with other members of the team. The TDS group was also more likely to indicate that the training improved their ability to understand the time it takes to move and execute various tactics on the battlefield. In contrast, the TDG group provided higher ratings on items pertaining to understanding the commander's intent and envisioning how situations could play out in the future.

Students reported that the TDS evaluated in this study was better than the TDGs in its ability to give students a better understanding of terrain and its importance in battle. Students also felt that while TDGs were good for planning, the TDS allowed them to test the plan and see the results of their actions. As several students noted, "CCM will actually show troops getting killed from your mistakes; TDGs don't."

Table 2. Student Survey Data Means (Standard Deviations)

Item	TDG	TDS	Significance (p)
2. [TDG/CCM] was often boring.	3.86 (1.07)	2.29 (1.25)	.027
7. The facilitator(s) of [TDG/CCM] were outstanding.	3.00 (1.00)	4.00 (0.82)	.063
9. I would like to participate in more [TDG/CCM] sessions in the future.	2.86 (1.07)	4.43 (0.79)	.009
11. [TDG/CCM] stimulated student interaction.	3.43 (0.54)	4.14 (0.90)	.096
19. The use of [TDG/CCM] has improved my ability to coordinate with other members of a team.	2.86 (1.07)	4.00 (0.58)	.029
21. Using [TDG/CCM] has improved my ability to consider how my mission relates to company's mission.	4.14 (0.38)	3.71 (0.49)	.091
26. Using [TDG/CCM] has improved my ability to understand the time it takes to move and execute various tactics on the battlefield.	2.43 (0.98)	4.00 (0.00)	.001
28. Using [TDG/CCM] has helped me develop strategies for managing the uncertainty of experiencing these events in combat situations.	3.14 (0.90)	3.86 (0.39)	.077
31. I believe the [TDG/CCM] experience will help me envision how similar situations could play out in the future.	4.29 (0.49)	3.86 (0.39)	.091

Instructor Survey

Three instructors completed the instructor survey. The instructors had 17 (one instructor) or 20 (two instructors) years of military experience. They had been at the Advanced School of Infantry for 2-4 years and had been instructing for 4-15 years. Because only three instructors completed surveys, statistical tests were not conducted on their rating responses. Comment data are presented in this section to summarize instructor opinions.

The instructors viewed the TDS as providing several advantages over TDGs. Like the students, the instructors appreciated the "realistic loss of combat effectiveness [as a result of] the loss of troops" and a "more realistic view of the battlefield." They also saw as a benefit the ability to replay the mission with different tactics and see the results of those changes. Finally, they noted its ability to motivate students to learn and to provide a realistic time environment in which students needed to make real-time decisions.

However, instructors also identified some disadvantages of the TDS in relation to traditional TDGs. First, the TDS takes more time and effort to conduct. Second, the TDS can get in the way of reality; students tended to forget that the people dying in the simulation are representative of real men. Students seemed to lose focus on the learning objectives, because "they just wanted to kill." Third, there was a steep learning curve in the TDS environment. After three sessions, students were still asking for technical help.

All three instructors felt that TDSs should not replace TDGs. One instructor summed up this opinion: "I think the best way to conduct training is brief the plan, assign the task, and scheme of maneuver on the sand table, then go in and fight it on CCM, then critique. They both complement each other."

Scenario-Based Cognitive Assessment

The pre- and post-test responses for the scenario-based cognitive assessment were scored by an SME blind to both the student's group and to the pre- or post-session condition. Student scores were analyzed with two-tailed independent samples t-tests. No significant differences were found between groups on any variables in the pre-test, indicating that any differences found on the post-test were due to the training intervention rather than differences in experience levels of the participants.

On the post-test, students in the TDG group were significantly more proficient at considering the operational effects of terrain when developing a plan than students in the TDS group (TDG M = 2.43, SD = .38; TDS M = 1.71, SD = .49, $p < .05$). This is an especially interesting finding as the students indicated that the TDS produced a better understanding of terrain than the TDGs. There was also a trend for students in the TDG group to be more proficient in modeling a thinking enemy (TDG M = 2.14 SD = .38; TDS M = 1.71, SD = .49, $p = .09$).

A trend for both the TDG and TDS groups to decline was observed as well, ($p = .07$) in overall performance from the pre-test (TDG M = 11.57, SD = 1.72; TDS M = 10.14, SD = 2.61) to the post-test. (TDG M = 9.71 SD = 2.56; TDS M = 9.0, SD = 1.63), perhaps due to participant fatigue. Datta (2000) found that giving subjects a wide array of outcome measures tended to produce uneven results due to fatigue and other outlying factors. Students in the present evaluation were asked to complete a number of different measures after a long day of training. Asking students to complete fewer measures would probably negate this discouraging trend.

Course of Action (COA) Exercise

The COA exercise scores were analyzed using two-tailed independent samples t-tests. Data from each of the three days were analyzed separately because each session involved a different scenario.

For one session, significant differences were found in plan quality between the TDG and TDS groups. No significant differences were found between the two groups in the pre-planning stage, confirming that the results were not the effect of differing experience levels. However, in the post-planning stage the TDG group did a significantly better job of providing sound reasoning behind their proposed course of action ($t = 2.25$, $p < .05$). In addition, the TDG group's predictions of the next 20-30 minutes indicated that they were significantly better than the TDS group ($t = 2.33$, $p < .05$) at understanding how the situation presented in the training exercise could reasonably progress. The third session produced no significant results between the two groups. This may have been the result of subject fatigue or ceiling effects.

Post-Exercise Knowledge Test

Mean scores for the TDS and TDG groups were compared on the post-exercise knowledge tests. A trend ($p = .08$) was found for the Patrol knowledge test only, favoring the TDG group (M = 3.9, SD =.38) over the TDS group (M = 3.14, SD =.90). An examination of the items indicated that the TDG trainees were better able to identify important terrain features represented in the scenario.

Multiple Choice Knowledge Test

The average number correct on the multiple-choice final exam was 64.0 (SD = 4.58) and 67.5 (SD = 7.0) for the TDG and TDS groups, respectively; a difference that was not significant using an independent samples t-test ($p > .09$).

Combat Orders Test

Only the patrol order was obtained from the combat orders test. The ratings pertained to the Execution section of a combat order and specifically to the quality of the order's representation of commander's intent, scheme of maneuver, fire support plan, tasks, and coordinating instructions. None of the five items discriminated the TDS from the TDG groups. Reasons may include no actual differences between the groups, lack of sensitivity of this measure (possibly because this was the third combat order produced by the groups and ceiling effects were obtained), or lack of reliability. A Cronbach's alpha performed was found to be .88, suggesting good reliability, although this may be inflated because of rater bias or halo.

Classroom Observation

The classroom observation data were coded using a BARS. Each session was divided into interactions, and the interactions were coded with the relevant BARS dimension or dimensions and a rating from 1 to 5 (1=novice; 5=expert) to indicate the maturity of tactical thinking displayed in that interaction. The BARS used was originally developed for planning only. However, during coding the TDS data yielded patterns where interactions seemed to fit under a particular BARS dimension and rating, but had no matching description for execution. Therefore, we developed definitions of what those ratings meant for execution and integrated them into the coding, then rechecked all codes to ensure we used the same standard for all the data.

The three most frequent BARS dimensions observed in the TDG data were "Know and Use All Assets" (Assets), "Visualize the Battlefield" (Visualize), and "Keep a Focus on the Mission and Higher's Intent" (Mission). These accounted for 53.8%, 13.5%, and 9.4% of the coded interactions out of 171 total. "Model a Thinking Enemy" (Enemy) was the next most frequent code, accounting for 7.6% of all interactions. The three most frequent BARS dimensions in the TDS data were Assets, Enemy, and Visualize, accounting for 63.8%, 17.3, and 9.1% of the total. Table 3 summarizes the BARS frequencies and percentages for both groups.

There were roughly twice as many coded interactions for the TDS sessions as for the TDG sessions, due to the format of the TDS. Students in the TDS group were split into two teams that operated against each other (i.e., force on force). Note that in analyzing the observational data, we coded interactions from all three training sessions even though the groups received different scenarios in the first session. These first session scenarios were similar, and our objective for the analysis of observational data was to assess the nature and quality of the interactions during the sessions, rather than performance outcomes as a result of the sessions.

DISCUSSION

The purpose of this effort was to identify strengths and weaknesses of a TDS compared to a TDG. Below, we discuss major findings from the evaluation. One important standard against which to judge the usefulness of these conclusions is the quality of the data set. The data obtained in the present evaluation represent a small number of participants (7 students per group), a factor which greatly reduces statistical power. In addition, many of the measures used were developed or modified specifically for this evaluation, and thus little is known about their reliability, sensitivity, or validity. Below we discuss several conclusions that seem warranted from the results and trends that we hope will guide future research.

Macrocognition

As explained in the introduction, macrocognition describes the emergent cognitive functions and processes that arise in naturalistic settings. Table 4 summarizes the preliminary indications of the relative advantages of the training formats for addressing the macrocognitive activities identified as most relevant for this evaluation. These are discussed below.

Coordination

The student survey data suggested that the TDS supported practice of team coordination to a greater extent than the TDG. In addition, the researchers observed a great deal of coordination between students in the TDS sessions, as students role-played different team members who were required to work as a unit to defeat the opposition. Team coordination was not a requirement, nor was it encouraged, in the TDG condition.

Mental Simulation

The COA exercise data indicated that in at least one training session, the students in the TDG group were better at predicting how a situation would evolve, suggesting that their mental simulation skills were superior to those in the TDS group. Likewise, a trend in the student survey data suggested that the TDGs helped students envision how similar situations could play out in the future more than the TDS, which in turn suggests that the TDG format promotes the exercise of mental skills to a greater extent than the TDS. It seems that all in all, the TDG format may better address mental simulation than TDS. However, it is also possible that a deepening of the cognitive content of the TDS, coupled with appropriate facilitation techniques, could negate any differences related to the development of mental simulation skills.

Uncertainty Management

A trend in the student survey data suggests the TDS helped students develop strategies for managing battlefield uncertainty better than TDGs. Uncertainty management is a key issue in military command and control; uncertainty is ubiquitous in any battlefield situation and competent management of it is required for successful leadership (Schmitt & Klein, 1996). We believe that given a larger sample size, we may have seen the TDS format producing a significantly greater impact on uncertainty management, as reported by the students, than the TDG format.

Planning

Written student comments suggested that the TDG encouraged more planning practice than the TDS. The student survey data indicated that students in the TDG group were better than those in the TDS group at considering the relation of their mission to the company's mission when planning the operation. In addition, the COA exercise data yielded a significant difference between the groups in providing the reason behind their COAs, suggesting that the TDG format might help students do a better job of thinking through the rationale for their plan.

Table 3. Observation Data

BARS	TDG Group		TDS Group	
Dimension	**Frequency**	**Percentage**	**Frequency**	**Percentage**
Assets	92	53.8%	203	63.8%
Visualize	23	13.5%	29	9.1%
Mission	16	9.4%	11	3.5%
Enemy	13	7.6%	55	17.3%
Contingencies	13	7.6%	2	.6%
Terrain	8	4.7%	15	4.7%
Big Picture	6	3.5%	1	.3%
Timing	0	0%	2	.6%

Planning
Written student comments suggested that the TDG encouraged more planning practice than the TDS. The student survey data indicated that students in the TDG group were better than those in the TDS group at considering the relation of their mission to the company's mission when planning the operation. In addition, the COA exercise data yielded a significant difference between the groups in providing the reason behind their COAs, suggesting that the TDG format might help students do a better job of thinking through the rationale for their plan.

Decision Making
Some students reported that TDGs provided more decision-making practice than TDSs. However, this finding seems counterintuitive. It may be that the formats impact decision making at different levels, with TDS enhancing basic tactical decisions—such as how to set up a base of fire or how to combine indirect mortar fire with movement to contact—and TDGs enhancing higher order tactical decisions, or even operational decisions such as how to defeat the enemy without destroying assets such as bridges or airfields that can support friendly operations at a later time.

Leverage Points
Leverage points are opportunities for making critical changes at a relatively low effort and a means by which COAs are generated in a problem-solving situation. Students' written comments suggested that TDGs allowed them to see that there was more than one way to accomplish a task. Envisioning a wide range of approaches for mission accomplishment is related to the facility of spotting leverage points and turning them into courses of action. Exposure to a wide variety of solutions to a problem, and the rationale for employing those solutions, is also a critical step in helping students decipher the subtle nuances in a situation that may lead to selection of certain COAs over others in a real situation.

Tactical Thinking and Mental Models
The eight dimensions identified in the BARS instrument capture the vital components of tactical thinking skills. The trends below are an initial conceptualization of the relative advantages of the two training formats, but each format should be further examined to produce more definitive results.

Focus on Mission and Higher's Intent
Student survey data showed a trend toward TDGs better helping students consider how their mission relates to the higher headquarters' (in this case, company) mission. Students in the TDG sessions considered the higher mission more than twice as often as the TDS students, and had a greater occurrence of higher ratings. The greater frequency of ratings and higher ratings in the TDG group indicate that these students thought about higher headquarters' mission more often and with a better understanding than those in the TDS group.

Model a Thinking Enemy
The TDG group considered the enemy 10% more than the TDS group. However, ratings of the observation data were higher on average in the TDG group, suggesting TDG students might be operating at a more advanced level when they were considering the enemy.

Consider Effects of Terrain
The issue of which training format better facilitates thinking about terrain is not easily resolved by these data. The scenario-based cognitive assessment data indicated the TDG group considered the operational effects of terrain more proficiently when developing a plan. In addition, the post-exercise knowledge test revealed a trend indicating the TDG group was better able to identify specific terrain features important to the patrol scenario. In contrast, the student surveys suggested students in the TDS group felt they had a better understanding of terrain. The observational data found the TDS and TDG groups thought about terrain

equally often, but not very much relative to the other dimensions.

Know and Use All Assets Available
The assets category was by far the most prominent dimension in the observational data for both groups. It accounted for 10% more coded interactions in the TDS group than in the TDG group. In the execution-based TDS, students continually manipulated teams and weapons in response to the changing dynamics of the simulation. In the TDG group, the discussions focused on various approaches for utilizing assets in order to accomplish the mission.

Consider Timing
Student surveys suggested that the TDS group developed a better sense of the time it takes to move units and execute tactics on the battlefield. This is no surprise as TDS builds in fairly accurate time-distance and time-on-target relationships, whereas TDGs are not executed in real time. Observational data, however, only coded two instances of timing consideration in TDS and none at all in the TDGs. It is likely that the learning that occurs relative to timing would not be evident in the communications between students, which is the basis of the observational data. Instead, this type of learning is probably internalized as students watch the progression of battlefield events and the time it takes their units to have an impact on the battlefield.

See the Big Picture
The TDG group considered the big picture more frequently than students in the TDS group, but neither group discussed it much. It may be that the particular scenarios used in this study did not require consideration of the larger picture in order for students to be successful.

Consider Contingencies and Remain Flexible
The TDS group hardly considered contingencies at all, whereas the TDG group considered them somewhat more often. We would have expected the TDS group to be faced with a greater need for contingency thinking than the TDG group, since the TDS requires students to adapt their approach continuously in response to an intelligent and dynamic adversary. It may be that students in the TDS group did not communicate about the contingencies that they were considering. Another possibility is that interactions that resulted from an individual's contingency thinking were communicated to team members in the context of asset utilization.

Visualize the Battlefield
Observation data showed the TDG group exhibited battlefield visualization somewhat more frequently than students in the TDS group. Additionally, the TDG group received higher ratings than the TDS group, suggesting that the TDG format not only encourages more visualization, but that the visualization occurs at a more advanced level than in the TDS format.

Table 4. Preliminary Indications of the Relative Advantages of TDS versus TDG Training Formats for Addressing Macrocognitive Functions

Macrocognitive Activity	Training Format Most Likely to Address the Macrocognitive Activity
Coordination	TDS
Mental Simulation	TDG
Uncertainty Management	TDS
Planning	TDG
Decision Making	Unknown
Leverage Points	Unknown

CONCLUSION

TDSs represent a new and potentially powerful paradigm in training and education in the military. Many of the training tasks for which simulation and live training have traditionally been used might be better accomplished by digital game-based trainers, which in many cases represent a more affordable alternative. The purpose of the present evaluation was to begin to assess the relative strengths and weaknesses of TDSs compared to TDGs, a pervasive training approach. Because of the lack of demonstrated power in the measures employed, our conclusions must be tentative. However, they provide guidance for future research. Our conclusions are summarized below.

- While there is some overlap, the TDG and TDS interventions evaluated in this research provide complementary training approaches. The TDG intervention enhanced mental simulation, planning, focusing on higher's intent and command level decision making to a greater extent than the TDS, while the TDS provided more emphasis on the timing of operations and team coordination.
- Trainees and instructors emphasized the utility of combining these training approaches so that planning would be supported by the TDG and execution supported by the TDS.
- TDSs possess inherent motivational properties.

- Some trainees in the TDS group were still struggling with the interface at the end of the evaluation period. Care must be taken in the design of TDS interfaces so that they require as little learning as possible.
- As TDSs are implemented, instructor support will be needed. Instructor support extends to exercise control and assessment of student performance. In addition, instructors will need support to fully integrate TDSs into training programs or courses.
- Finally, our overall evaluation strategy appears to be sound; our skill taxonomy allowed diagnostic assessment of the relative strengths and weaknesses of TDGs and TDSs, although the psychometric properties of the measures still need to be assessed and documented. This type of approach should be continued in future evaluations, with greater coverage of the skill taxonomy and more attention devoted to skill-based assessments. The assessment of team coordination is another area that should be enhanced in future TDS evaluations.

The results of this pilot study, preliminary as they are, have led to critical insights that will serve to influence the design and development of a second generation of TDSs to train critical thinking skills. For example, next generation USMC TDSs will have improved capabilities to view terrain such as a zoom feature and the ability to observe terrain anywhere from a 60 deg to 90 deg angle. Scenario editors and triggering devices (e.g., scenario events; static and moving models) are being developed for use by instructors to focus students on specific learning objectives. Improved scenario editor and trigger functions will aid in future data collection efforts to (1) verify the effectiveness of TDSs as training tools in comparison to an existing classroom standard, and (2) identify and test instructional approaches that provide optimal learning to inform schoolhouses on how best to use TDSs to train specific tasks.

ACKNOWLEDGMENTS

We would like to acknowledge the support of the instructors and students in the U. S. Marine Corps Infantry Platoon Sergeant Course taught at the Advanced Infantry Training School at Camp Geiger, North Carolina. They allowed us total access to their course and generously gave their time to complete the data collection measures.

REFERENCES

Datta, L. E. (2000). Seriously seeking fairness: Strategies for crafting non-partisan evaluations in a partisan world. *American Journal of Evaluation*, 1-13.

Deckert, J. C., Entin, E. B., Entin, E. E., MacMillan, J., & Serfaty, D. (1994). *Military decision-making expertise: Final report*. Fort Leavenworth, KS: Army Research Institute.

Kirkpatrick, D.L. (1976). Evaluation of Training. In. R.L. Craig (Ed.), *Training and Development Handbook*. NY: McGraw-Hill Book Company.

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems, 18*(3), 81-85.

Kraiger, K., Ford, J.K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.

Lussier, J. W., Ross, K. G., & Mayes, R. (2000). Coaching Techniques for Adaptive Thinking. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference.* Orlando, Florida.

Lussier, J. W., Shadrick, S. B., & Prevou, M. I. (2003). *Think like a commander prototype: Instructor's guide to adaptive thinking* (No. ARI Research Product 2003-01). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Muchinsky, P. M. (2003). *Psychology applied to work*. Melmont, CA: Wadsworth/Thomson.

Riggio, R. E. (2000). *Introduction to industrial/organizational psychology*. Upper Saddle River, NJ: Prentice Hall.

Ross, K. G., & Lussier, J. W. (1999). A Training Solution for Adaptive Battlefield Performance. *Proceedings of the Interservice/ Industry Training, Simulation and Education Conference.* Orlando, Florida.

Schmitt, J. F., & Klein, G. (1996). Fighting in the fog: Dealing with battlefield uncertainty. *Marine Corps Gazette, 80,* 62-69.

Shadrick, S. B., & Lussier, J. W. (2002). The Application of Think Like A Commander in the Armor Captains Career Course. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference*. Orlando, FL.