# AUTOMATED SOCRATIC TUTORS FOR HIGH-LEVEL COMMAND SKILLS

**Eric A. Domeshek**
**Stottler Henke Associates, Inc.**
**Arlington, MA**

**Elias Holman**
**Stottler Henke Associates, Inc.**
**Arlington, MA**

**Karol G. Ross**
**Klein Associates, Inc.**
**Fairborn, OH**

**Abstract**

Under stress, human decision-makers revert to their best-practiced habits. This includes military commanders who may fail to act effectively under pressure for lack of sufficient practice. The US Army Research Institute (ARI) developed a training methodology emphasizing repeated exposure to small challenging vignettes enabling drill on command decision-making. Rather than role-play to a simulated conclusion, mentors focus on analyses and dialogues that explore reasoning and rationale. Adoption of this methodology in courses at Forts Leavenworth and Knox is helping to validate this approach to honing command skills. However, intense practice with human mentors is problematic as there are generally too few expert mentors available.

This paper describes an ongoing Small Business Innovative Research (SBIR) project to develop automated intelligent tutors to fill the role of expert mentors. This project is developing a novel capability to understand, critique, and discuss proposed courses of action in a Socratic mode, guiding the student as an expert would. The approach emphasizes multi-modal interaction (e.g., language and graphics), models of expert human tutors, and development of authoring tools to reduce training system costs.

The paper presents data from analyses of expert mentor and student dialogues during "tactical decision games." It then describes how this data is used to develop and assess the project's intelligent tutor. Additional preliminary data from early informal formative evaluation of the Phase I prototype system, and initial student feedback on some Phase II refinements is also reported.

Ongoing complementary efforts include a related Phase II SBIR with a different ITS approach, and a computer-based program developed by ARI that human instructors are using in the Armor Captains Course at Fort Knox's University of Mounted Warfare.

**Author's Biographies**

Dr. Eric A. Domeshek is an Artificial Intelligence (AI) Project Manager at Stottler Henke Associates, Inc. He received his Ph.D. in Computer Science from Yale University, where his work focused on cognitive modeling and technology, most especially on development of Case Based Reasoning (CBR). While working as Research Faculty at the Georgia Institute of Technology, he helped launch the EduTech institute, and became involved in educational applications of AI and CBR. He continued to work on educational and training technology while on faculty at Northwestern University's Institute for the Learning Sciences. For the last three years, Dr. Domeshek has conceived and managed a variety of Intelligent Tutoring System (ITS) projects at Stottler Henke. His main effort currently focuses on dialogue-oriented tutors such as the ComMentor system described here.

Mr. Elias Holman is an Artificial Intelligence Software Engineer at Stottler Henke Associates, Inc. He received his BA in Music Technology from Oberlin College, and will start in the fall of 2002 at the Harvard School of Education in the Technology in Education program. Over the last two years, Mr. Holman has worked on several ITS projects at Stottler Henke, as well as projects focused on web-based collaboration.

Dr. Karol G. Ross is a Senior Research Associate at Klein Associates Inc. She is currently pursuing research and development of decision skills training for both individuals and teams in the areas of battlefield tactical thinking and firefighting and the development of training for military situation awareness skills. Formerly, as a Research Psychologist for the U.S. Army Research Laboratory, she conducted research and training development for adaptive tactical thinking training at the Brigade level and development of PC-based simulation training for Field Artillery decision making. She also previously served as a Senior Researcher for BDM International studying battle staff performance and as a Research Psychologist for the Army Research Institute in Heidelberg, Germany 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.

# AUTOMATED SOCRATIC TUTORS FOR HIGH-LEVEL COMMAND SKILLS

**Eric A. Domeshek**
**Stottler Henke Associates, Inc.**
**Arlington, MA**

**Elias Holman**
**Stottler Henke Associates, Inc.**
**Arlington, MA**

**Karol G. Ross**
**Klein Associates, Inc.**
**Fairborn, OH**

## INTRODUCTION

Military commanders must make complex tactical decisions in the heat of battle. While they spend years studying and training for such situations, their opportunities to drill on the components of tactical decision-making skills are actually quite limited. Much of the training that is available is too coarse-grained and infrequent. A key example of battle command training is unit rotations to the National Training Center (NTC). However, any given officer is likely to visit NTC only a few times during their career; when there, they spend days planning and executing with very little incremental coaching and feedback. Most forms of simulation-based training share these defects, though of course to much lesser extents.

Low frequency practice with little feedback is inadequate because we know that human decision-makers revert to their best-practiced habits when under stress. The US Army Research Institute (ARI) is attempting to address the question of how a high volume of deliberately structured practice (Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson, 1996) can be provided to commanders so that the component skills of effective tactical decision-making are firmly ingrained. To that end, they have developed a training methodology emphasizing repeated exposure to small challenging vignettes, enabling drill on command decision-making. Rather than role-play to a simulated conclusion, mentors focus on analyses and dialogues that explore reasoning and rationale. Adoption of this methodology in courses at Forts Leavenworth (Lussier, Ross, & Mayes, 2000; Ross, & Lussier, 1999) and Knox (Shadrick & Lussier, in preparation) is helping to validate this approach to honing battle command thinking skills.

This approach is, in many ways, similar to the spirit underlying the training formats known as Tactical Decision Games (TDGs) or Quick Decision Exercises (QDXs) as practiced, for instance, by the Marines and Army respectively. The differences primarily derive from the amount of mentoring and interactive support offered to the student, and the extent to which that mentoring is focused on a set of key command themes or cognitive skills derived from the results of extensive interviews with expert tactical thinkers (Deckert, Entin, Entin, MacMillan, & Serfaty, 1994; U.S. Army Research Institute, 2001):

1. Focus on mission and commander's intent
2. Model a thinking enemy
3. Consider the big picture
4. Consider effects of terrain
5. Make effective use of all resources
6. Prioritize and take action in time
7. Visualize a changing situation
8. Plan flexibly and for contingencies

However, any military training system that depends on intense practice with human mentors is problematic, as there are generally too few expert mentors available. In response, ARI is also supervising a pair of Small Business Innovative Research (SBIR) projects to develop automated Intelligent Tutoring Systems (ITSs) (Ong & Ramanchandran, 2000) to fill the role of expert mentors. The work reported here has been performed in the context of one of those projects. Stottler Henke Associates, Inc. and its partners at Klein Associates, Inc. are developing an ITS called ComMentor—an automated Commander's Mentor.

ComMentor will provide a novel capability to understand, critique, and discuss proposed courses of action in a Socratic mode, guiding the student as an expert would. Socratic instruction is a kind of teaching interaction typically applied in high-level professional education (e.g. law and business) and most often characterized by its external form: the teacher asks a lot of questions, and the student answers. These surface features actually reflect the deeper nature of Socratic instruction: it is an interactive and highly engaging form of scaffolded constructivist pedagogy. The question and answer format keeps the student engaged, but lets the teacher lead. The questions are posed in a sequence that leads the student to reconstruct the logic of expert situation analysis and decision-making for themselves. In ComMentor, our approach to the design of a Socratic tutor for battlefield command reasoning skills emphasizes multi-modal interaction (e.g., language and graphics), models of expert human tutors, and development of authoring tools to reduce training system costs.

We have identified Captains (specifically in the U.S. Army and U.S. Marine Corps) as the most promising grade of officers for which to develop this training initially. There are a large number of Captains in the services, and their access to senior officers who might be able to provide mentoring is typically limited. Yet their responsibilities in combat are quite large, especially when one considers the doctrinal requirement to think two echelons up from their position as company commanders or battalion staff members.

At the same time, we have aimed many of our scenarios at a level higher than that at which current Captains are normally expected to operate. For instance, some scenarios involve command of a battalion. Others involve command of a company in the Objective Force, with capabilities that demand tactical skills akin to what might normally be expected of a Colonel. One way to develop higher-level command skills is to be thrust into the position of commanding larger or more capable units. Since focusing on higher commanders' intent is one of the most important command reasoning skills, putting a student in the shoes of a higher-level commander is an especially valuable kind of training. In our observations, it also tends to have quite positive motivational impact on students.

### Research Focus for this Paper

Our ITS development efforts are being guided by extensive observations of active-duty military officers engaged in TDG training with recognized expert mentors. In Part I, this paper presents data and analyses of such tutoring sessions, and then describes how this data is used to drive design of the project's intelligent tutor. In Part II, we report on preliminary formative reviews of early prototype versions of ComMentor, including expert and student feedback.

## PART I: OBSERVATIONS OF HUMAN EXPERT MENTOR TUTORIAL SESSIONS

During the Phase I project and in the first third of the Phase II project, we have conducted a total of 30 TDG sessions with active duty officers from the US Marines and Army. This section describes those sessions and their results in detail.

### Objectives

Our tutorial observation sessions have been intended to provide critical information on a variety of issues:

1. Ranges of student ability
2. Expert mentoring techniques
3. Typical student behavior in specific scenarios
4. Typical mentor approaches and sequences in specific scenarios
5. Specific scenario contents and presentations
6. Relevant background domain knowledge

Data on these issues are being used to inform the design and implementation of the ComMentor system. In particular, ComMentor is conceived as a scenario-based system, and so in our Phase II work we have been using these sessions as a way to develop detailed understandings of a set of six scenarios intended to form the initial training corpus for the delivered system.

### Method

***Overview.*** Marine and Army Captains were invited to participate in individualized Tactical Decision Game training delivered in a mixed live/electronic format. These TDGs were designed and mentored by a pair of retired Marine officers recognized for their expertise in tactical training (see Schmitt, 1994; Schmitt, 1996). In all cases, one of the mentors was present in the room with the student while the other was connected by a voice phone line and a shared whiteboard application hosted in a web browser. The shared whiteboard was used to present task organization wire diagrams and scenario maps; it supported shared annotations, pointers, and screen capture. The on-site mentor provided technical support as well as normal mentoring (informed, on account of collocation with the student, by the interpersonal cues that might be missed by the remote mentor).

Most of the TDG sessions lasted approximately 1.5 hours, except for the final sessions devoted to scenarios based on a futuristic Objective Force, which required over 2 hours because of time spent discussing the projected systems and capabilities envisioned in Future Combat Systems (FCS) concepts. Sessions started with an informal interview bearing on the student's background. Then the project, team, and purpose of our session were introduced. The mentors took the student through an introduction to the web-based whiteboard tools, and then finally, introduced the scenario for the session. At every point, students were allowed and encouraged to ask questions.

Sessions were tape-recorded, and occasional whiteboard screen-shots were taken during the interaction to capture the state of a shared map with its unit positions and other annotations. Sessions were also generally observed by at least two project team members in addition to the two mentors. In most cases, those observers were not present in the room, but were also connected to the teleconference and the shared whiteboard. One project team member was skilled in cognitive task analysis and modeling, while the other represented AI model design and implementation expertise. As noted earlier, the main purpose of these observation sessions was to provide data on which to base a cognitive model of battlefield command reasoning skills tutoring, preparatory to embedding that model in a functioning computer tutor.

***Students.*** Students were active duty Marine or Army Captains. Marines were primarily tactics instructors recruited from the Quantico training center. The Army officers were drawn from a range of units and specialties at Fort Riley. A total of 17 Marine Corps Captains participated in the thirty tutorial sessions. Four Marine Corps officers participated in two different sessions. One participated in three sessions over the course of the study. Those who participated in more than one session did not repeat the same scenario. The Marine Corps officer, all affiliated with combat arms, had a range of experiences, but all had been Platoon Leaders and Company Executive Officers. Their length of service varied from 6 to 14 years. Eight U.S. Army Captains participated in the tutorial sessions. Of those, two were currently Company Commander. The Army officers were affiliated with combat arms and other branches of the service to include Intelligence, Signal, Armor, and the Quartermaster Corps.

***Materials.*** The primary materials in this study were the TDG scenarios themselves. During Phase I, we held six sessions and worked with three different scenarios. During Phase II, we held a total of twenty-four sessions: four sessions for each of six different scenarios. Table 1 below summarizes the Phase II scenarios.

| Phase II Scenarios | Topic | Student Role |
|---|---|---|
| 1$^{st}$ 2 Scenarios | Present-day armor | Battalion Cmdr |
| 2$^{nd}$ 2 Scenarios | Interim Brigade | Battalion Cmdr |
| 3$^{rd}$ 2 Scenarios | Objective Force | Company Cmdr |

Table 1. Scenarios for Phase II Tutorial Observations.

Each scenario typically consisted of a narrative that was read to the student with some supplementary commentary by one of the mentors. In addition, when the friendly or enemy task organizations were complex enough (and relevant and well known enough) a wire-diagram of those organizations was generally provided. Finally, every scenario had a (generally abstracted) map of some restricted area of relevant terrain. Mentors also provided answers (either factual, or extemporary scenario extensions, as appropriate) to most questions asked by students.

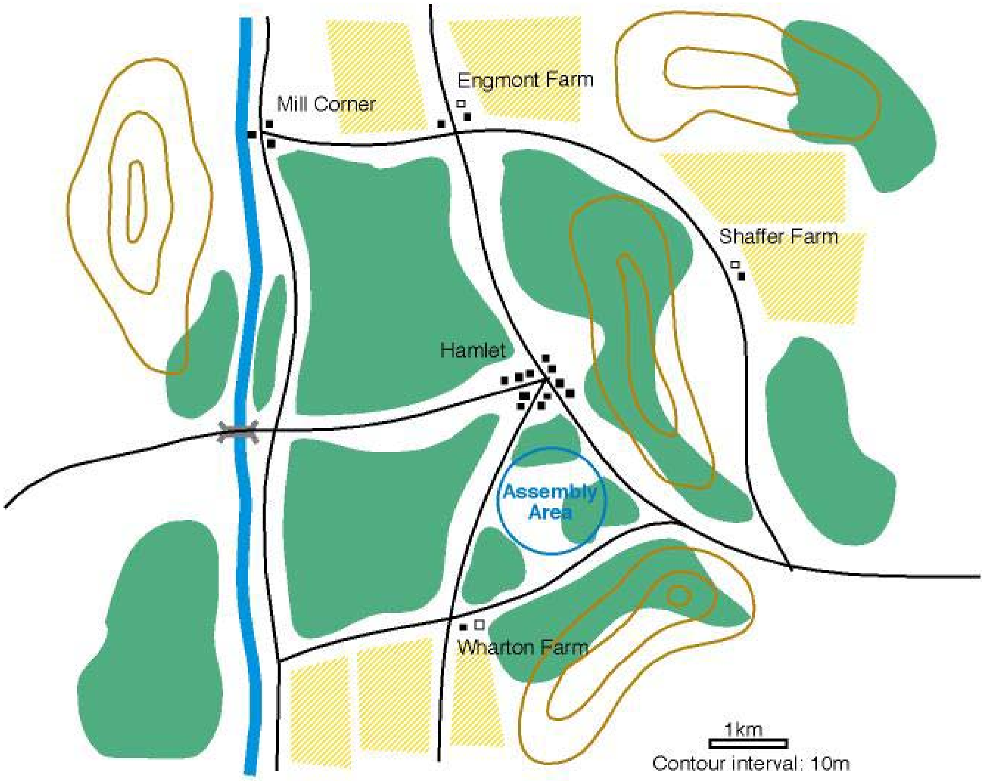


Figure 1. Map for the “Enemy Over the Bridge Scenario”.

**Enemy Over the Bridge**

You command a balanced tank-mech battalion task force consisting of 2 tank companies (A and B), two mech infantry companies (C and D), a scout platoon and a mortar platoon, plus your medical, support and maintenance platoons organized into a company train.

You are fighting a capable, mechanized enemy equipped with T-62s and BMPs and supported by towed and self-propelled artillery.

Host-nation forces hold the bridge and the river line to your west. You have been told the river is unfordable. Reconnaissance elements are operating west of the river. In 48 hours, the division begins a major offensive west across the river to destroy enemy forces in zone, with the main effort in your brigade's zone. Your battalion will spearhead the brigade's attack.

You have been instructed to occupy the assembly area shown on the map east of Hamlet in preparation for the 0500 attack the morning after next. You are moving to the assembly area as shown. At 0100 your scout platoon, which is forward reconnoitering the route to the assembly area, makes the following report:

> *Battalion, be advised have just made contact with a host-nation motorized reconnaissance patrol that was operating west of the river but about 2 hours ago was forced east across the river under fire. They came across the bridge and then via Wharton Farm. The reconnaissance patrol leader reports there is no sign of friendly forces holding the river line or the bridge and that enemy mech infantry and some tanks have been moving east across the bridge for almost 2 hours. He says he counted 10 T-62s in the last half hour; does not know how much mech. He says he has reported this twice to his higher headquarters. Over.*

A few minutes later the scout platoon leader adds the following:

> *Be advised we've got enemy mech infantry occupying our assembly area in strength. I say again they are enemy and not host-nation forces. I've got a solid visual on several BMPs. Don't know the size, but I estimate at least a company. They seem to be still moving into the area, over.*

*Suddenly, you start to see artillery impacting in the woods just north of Alpha Company at the head of the battalion column.*

Moments later, you hear automatic weapons fire from the direction of the assembly area. *"We're in contact!"* the scout platoon commander shouts over the radio.

What do you do?

Figure 2. Narrative Introduction to the "Enemy Over the Bridge" Scenario.

As a sample, Figure 2 shows the narrative introduction to first scenario used during the Phase II sessions. Figure 1 shows the accompanying map. In this case, the Blue task organization was considered simple enough that no wire diagram was provided. Likewise, since the Red organization was unknown, no diagram was provided for that force either. Note that when the narrative says, "You are moving into the assembly area as shown." a set of movable icons were positioned on the map to indicate roughly where the student's forces were supposed to be located at that time. Both the student and mentors were free to later move those icons as the dialogues about the student's response unfolded.

***Procedure.*** As described, each session started with introductions of the student, background on the project, participants, purposes, and process of the session, and a tutorial on the web-based tools. When all that was done, the TDG proper began with the reading of the scenario narrative, and proceeded through a series of stages, as reported in the Results session below.

When the mentored scenario itself was over, each session ended with a brief post-scenario interview. In later runs this centered on the student's prior tactical training experiences, their appreciation of this experience, and their potential interest in additional sessions, either with live mentors or with a computerized mentor.

## Results

We conducted these sessions with the objective enumerated above. Here we discuss our observations bearing on those objectives.

***Ranges of Student Ability.*** Captains are a remarkably diverse lot. They can range in experience from 6 years to 14 years of active duty service. They can be deeply immersed in tactical art, or they can be extremely remote from it. As an example of one extreme, we worked with a very experienced Captain (an armor officer, with almost a decade as an enlisted soldier, and experience in Desert Storm, before even receiving his commission) who was extremely quick, both to act appropriately, and to grasp the flaws in his approach as soon as they were even hinted at. At the other extreme, one participant, lacking a combat arms background, also seemed to lack the language and mental models to grasp the variety of the elements in the tactical situation, and essentially froze when first presented with a TDG situation description.

Of course we primarily saw students distributed across the broad intermediate range of background and ability. Sampling the full range was quite valuable to our work, as it ensured that we got a more complete picture of how students might respond to the TDG scenarios. It also drove home the point that there will be limits on the final system's range of applicability; some students—especially those with too little experience—will not be well served by the tutorial behavior authored to address the needs of the more average Captain.

***Expert Mentoring Techniques.*** One of the major purposes of these observations was to see how expert mentors conduct the kind of training we are aiming to simulate in the ComMentor ITS. Accordingly, we report here the general mentored TDG interaction structure observed across our many sessions as a *Result* (rather than detailing it in the *Procedure* section above).

The scenario's introductory narrative generally ended with a challenge to the student to respond to the situation. In some scenarios, students were invited to start responding as soon as they wanted to (that is, they did not have to wait until the end of the narrative). In complex scenarios, students were asked to perform an orienting task, such as a hasty terrain analysis. That orienting activity might be requested in the middle of the narrative, in scenarios where a crisis event potentially called for immediate action.

In all scenarios, a degree of time pressure was created by setting a limit on how long the student could think before acting. In some scenarios—particularly those with an unfolding crisis—the mentors reported consequences if the student delayed (e.g. changes in the situation, usually not to their advantage).

The students almost always produced some kind of response to the situation, though often the mentors had to help in drawing it out. Often there was a mini role-play with the student issuing fragmentary orders, as over a radio net, while a mentor played the parts of subordinates trying to understand and react to the orders (or superiors responding to reports and requests).

Once a student felt their response was complete, the mentors would typically proceed to probe for their rationale: what did they think was going on in the (usually ambiguous) situation? What were they trying to accomplish and why?

These rationale dialogues generally transitioned into dialogues bearing on the reasoning themes central to our training objectives: What did they think the enemy was trying to do and why? What did they think their higher commander would want them to do? What did they think would happen in some amount of time if they adopted their proposed course of action? Dialogues, steered by the mentors, ranged from general themes, to specific facts of "battlefield calculus," to detailed critiques of particular courses of action, and proposals for alternatives.

Each scenario ended with several minutes of mentor-led reflection. The mentors consistently asked a suite of meta-level questions: How do you think you did? What do you think you did best? What do you think you did worst? What did you learn?

***Typical Student Behavior in Specific Scenarios.*** By running each scenario several times (four times each for the Phase II scenarios) we were in a position to see a range of possible student behavior on each problem. The first observation was that—as suggested by our expert mentors in advance—there were definite patterns and frequent commonalities in students' responses. For instance, in the "Enemy Over the Bridge" scenario, the most common initial reaction was to deal forcefully with the firefight in the assembly area, but to essentially ignore (or treat as a lower priority) the continuing flow of enemy troops over the unsecured bridge. With common student response patterns, it follows that there were common topics in mentored dialogues.

The second observation was that different students required different amounts of dialogue and prompting (often forcing the mentors to try several different approaches) in order to recognize key points about the situation. Getting the student to see the importance of the bridge might take as little as asking "Can you identify any pieces of key terrain in this situation?" or as much as a long dialogue on the Blue and Red intentions in the situation, and explicit visualization exercises about what was likely to happen over time.

***Typical Mentor Approaches and Sequences in Specific Scenarios.*** As already noted, common student responses tended to elicit common mentored dialogue topics. We are currently in the process of inventorying the dialogue topics for each scenario by analyzing and comparing across session transcripts.

For instance, again drawing on "Enemy Over the Bridge," we can identify recurring dialogue topics such as: (1) getting the student to realize that they must stop the flow of new enemy forces over the bridge; (2) getting the student to realize that attacking through the assembly area, on to Hamlet, and then to the bridge is equivalent to attacking an enemy surface rather than a gap; (3) getting the student to enumerate a plausible set of alternate courses of action for controlling the bridge; and (4) getting the student to critique each of those courses of action on its strengths and weaknesses.

***Specific Scenario Contents and Presentations.*** For each scenario observed in Phase II we gathered the basic scenario materials from our mentor/authors for adaptation and incorporation in the system. For instance, the materials shown in Figure 1 and Figure 2 have been converted into a format appropriate for use in the ComMentor prototype (see, for instance, Figure 3 and Figure 4 below). We also gathered the force structure diagrams and relevant icon sets used by our mentors (again, as reflected in Figure 3 and Figure 4).

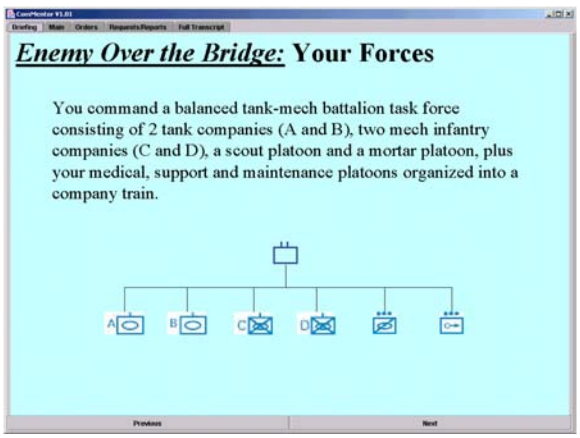


Figure 3. Introductory Briefing for "Enemy Over the Bridge Scenario" in ComMentor Prototype.

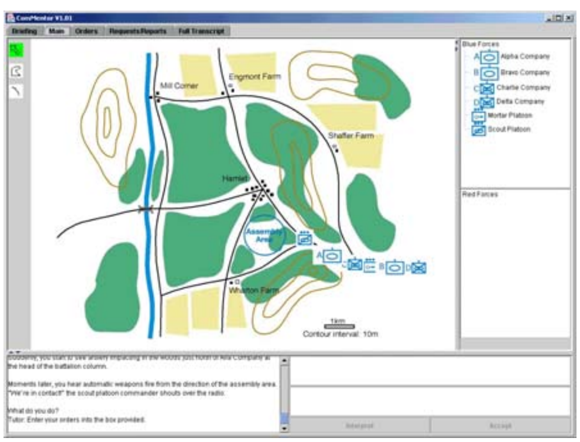


Figure 4. Main Screen in ComMentor Prototype Showing "Enemy Over the Bridge" Situation.

***Relevant Background Domain Knowledge.*** The expert mentoring we observed is well beyond the capability of any current AI approach *in the general case*. However, we do not necessarily need to match all aspects of our expert mentors' performance to have a useful system. Even for those aspects we do care about, an advantage of a scenario-centered approach is that we do not always need to build our system for the general case.

While ComMentor does not need to be able to reason as richly and creatively as our experts, it does need to be able to represent all of the kinds of statements of fact, intention, and conjecture that students typically utter during a TDG interaction. We can also reasonably aim to have it answer many basic questions of fact and simple computation that our experts classify as "battlefield calculus" (e.g. How many tanks in an enemy battalion? How fast can a tank company move over a paved road? How far and how fast, and with what probability of kill can a given tank shoot?).

The transcripts of our observed tutorial sessions provide an extremely rich source of information to address these domain knowledge representation issues. Combining analysis of transcripts and guidance from our expert mentors with other sources of AI representation for the military domain (e.g. results of the DARPA HPKB program) has already allowed us to assemble a rich ontology for relevant aspects of the military domain.

## Discussion

The point of holding these tutoring observation sessions was to gather data on what the ComMentor ITS should be able to do, and how it should be structured. Here we discuss some aspects of ComMentor's design that were determined by the study results.

***User Interface Structure.*** Several aspects of the TDG interactions we observed seem to be critical to capturing the flavor of these sessions and enabling the kind of deep engagement and constructivist learning these sessions engender. Here we enumerate several of the most important.

1. Situation map as the central visualization
2. Force structure diagrams
3. Dialogue-based interaction

Based on our observations and advice from our mentors, the ComMentor prototype devotes the bulk of its user interface to a map of the relevant territory, which serves as the central visualization for the scenario. While the terrain displayed can be limited, and the detail (purposefully) simplified, we try to ensure that the whole map is fully visible at all times (e.g. there is no need to scroll madly to see the entire area of interest). The map display is manipulable, both by the student and by the (automated) mentor, so it can reflect the evolution of the situation over time, and so that it can be used to express geographically related meanings that are not as easily stated in other modes, such as language (e.g. positions and relative positions, paths of movement, lines of visibility, etc).

The prototype's map is complemented by a pair of hierarchy displays that depict what is known of the Blue and Red task organizations. These displays serve both as mnemonics to help the student remember some of the key facts of the scenario, and also as pallets from which the student can pick icons to place on the map.

Finally, to complement these visualizations, the student's main interaction with the system is language-based—currently written (typed) text is expected to be the most frequent kind of student input and tutor output. As technology matures, spoken language input and output may become feasible. We have begun collecting data to see how typed text interaction differs from spoken language interaction.

Natural Language Processing (NLP) technology is not completely mature or robust, largely because, in the general case, it addresses an unconstrained problem. Of course, not overtly constraining the student is precisely what we hope to accomplish by adopting NLP for ComMentor. Not only would some alternate input system based on buttons, menus or utterance templates feel artificial, but we fear it would slow, inhibit, and lead the student too heavily, ultimately interfering with their focus on the scenario and their construction of knowledge (and running counter to the goals of Socratic interaction). We are hopeful that in the specific circumstances of ComMentor, NLP will prove tractable, and we have achieved promising preliminary results to date.

***Dialogue Structures.*** In order to adopt NLP successfully for a major piece of the system input, we have to find some way to restrict the interpretation problem. To this end, we note the following four observations:

1. Observed student/mentor dialogues were frequently built around the explicit reasoning themes ComMentor is being designed to teach;
2. Common student responses frequently elicited common dialogue topics;
3. The mentors, for the most part, remained in control of the course of the session, (e.g. choosing topics and framing the dialogue on those topics; but
4. Student's required different depth of detail in dialogues (or number of alternate dialogue approaches) to arrive at important conclusions.

ComMentor, then, was designed as a dialogue-oriented system where the tutor maintains the primary initiative in introducing and pursuing dialogue topics based on the evidence it is able to collect about the student's understandings and intentions. Some initiative may be temporarily ceded to the student in cases where they interpolate questions (primarily about session procedure or domain facts) into the flow of the conversation. But for the most part, the tutor asks the questions, and then tries to interpret the student's responses in relation to those questions; if that fails the tutor will try to interpret the input with respect to other aspects of the student state (e.g. the history of past dialogues, the stack of pending interactions, the agenda of planned dialogues).

For each scenario, dialogue topics are tied to evaluation patterns specifying student behaviors or rationales that, according to the expert scenario author, are worthy of critique or complement. These scenario-specific evaluations (e.g. the student says they are aiming to take control of the bridge, but they are doing so by attacking through the assembly area and Hamlet) are themselves specializations of domain-general evaluations tied to the system's curriculum (e.g. the principle of "surfaces" and "gaps"), which can in turn often be tied to the system's main themes of tactical reasoning (e.g. make effective use of resources).

We have so far identified four categories of dialogues:

1. **Set-Pieces:** These are mostly one-sided dialogues in which the tutor has something to say, and just says it. There may be some limited pauses for student input (e.g. confirmation, or opportunities for questions), but they do not materially affect the course of the presentation. Examples include initial scenario briefings, and "war-story" anecdotes intended to make a particular point.
2. **Static Fact Tree Dialogues:** These are tree-structured interactive dialogues, in which the goal is to get the student to see and acknowledge some fact about the situation that the tutor believes they may be missing. The tutor prompts (perhaps repeatedly in different ways) to get the student to state the key fact. If the student doesn't give the desired answer, then the system walks a level down the dialogue tree and tries to get the student to see/acknowledge a set of more specific facts that together constitute a good argument for the higher level target fact. Again, the system may have alternate arguments it can try in turn in its effort to get the student to reach the desired conclusion. Examples include arguments for why taking control of the bridge must be a top priority.
3. **Enumeration Dialogues:** At the top level, these are list-structured dialogues (however, each item in the top-level enumeration may have a static fact tree of arguments beneath it). The object is to get the student to generate (e.g. brainstorm) alternatives for some decision or situation interpretation point. Examples include enumerating courses of action that will achieve control of the bridge, or ways to maneuver the scouts out of the way of a battle at the assembly area.

4. **Pro/Con Dialogues:** At the top level, these dialogues are structured as two lists of arguments—those in favor of and those against a course of action or situation interpretation (again, each item at the top-level may have a static fact tree of argument beneath it). The object is to get the student to generate as many of the pro and con factors/arguments as possible. Examples include evaluating courses of action intended to achieve control of the bridge, or evaluating alternate orders that might be given to the scouts to get them out of the battle at the assembly area.

***Overall Session Control.*** A major ongoing aspect of our tutorial data evaluation is a study of tutor action types and the conditions that elicit them. Preliminary analyses have identified close to 100 reasonably distinct tutor move types. Based on their clustering, frequency, significance, and tractability, we have already begun to design many of these tutor actions into ComMentor.

Evaluation patterns are responsible for nominating dialogues that might be useful for a particular student. The tree and list structures of individual dialogues imply some degree of control over the course of the ComMentor interaction. But there are many other kinds of session control decisions that the tutor needs to be able to make: whether or not to actually pursue any particular dialogue, how to order the launch of queued dialogues, whether to abandon or interrupt pursuit of an active dialogue, when to pursue additional information that might help suggest useful dialogues, and so on.

Figure 5 summarizes the ComMentor processing cycle, which breaks down into three main blocks: Student **Input**, Tutorial **Planning**, and Tutor **Output**.

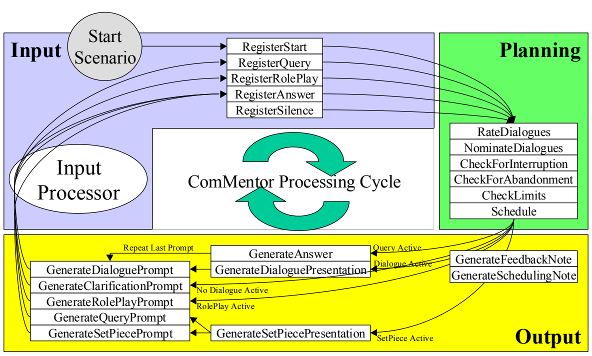


Figure 5. Diagram of ComMentor's Proposed Control Cycle.

The Input block shows a general Input Processor that is responsible for interpreting multimodal input from the student and registering that input into the three major student utterance categories of query (the student is asking a question), role-play (the student is taking some action in the context of the scenario), or answer (the student is responding to one of the tutor's questions).

Next up is the Control block, whose primary job is to maintain the dialogue agenda. It does that by scanning available evaluations to see if recent student input has changed the match-score of their patterns. It may nominate new dialogues (or disqualify old ones). Based on such changes, as well as the nature of the student input (e.g. query *vs.* answer) and even the passage of time, decisions may be made to disrupt the current dialogue flow (either a temporary interruption or a permanent abandonment of the dialogue in progress). Finally, the overall dialogue agenda may be updated and reordered based on ComMentor's shifted assessment of the student's beliefs.

Finally, the cycle closes with the Output block, which proceeds in three main steps: prefatory notes, main presentations, and closing prompts. Prefatory notes allow for insertion of discourse markers such as tutor feedback (encouragement, commiseration), and explicit acknowledgment of scheduling changes (interruption, abandonment, deferral, timeout). A main presentation is generally either a piece of a dialogue (including part of a set-piece) or an answer to a student question. Finally, the closing prompt asks the student to provide some particular kind of follow-up input, intended to bias the student towards providing the kind of input the system is then prepared to understand as input.

## PART II: PRELIMINARY PROTOTYPE FORMATIVE EVALUATIONS

Formative evaluations are being conducted on a regular basis during the Phase II project. So far we have conducted an expert review of the Phase I ComMentor prototype, and gathered a first round of expert and student feedback on early Phase II refinements to the ITS. This section briefly describes the initial and exploratory findings from those evaluations.

### Questions and Methods

The questions we seek to address include the following:

1. Does the system offer an appropriate kind of interaction, and supporting visualizations?
2. Is the interface design and layout clear?
3. How close is the prototype to being able to handle the kinds of input students produce?

Early in the Phase II project, members of our extended project team and interested parties from ARI were shown screen shots and live screens from a slightly updated Phase I ComMentor prototype. These were used as the basis for discussion of system design and possible features for Phase II. In June of 2002, we had another such review, based on a now somewhat more operational system. The system evaluators have included our two expert mentors (our Subject Matter

Experts, or SMEs), four staff members from Klein Associates, and four members of the ARI Ft. Knox unit. In addition, a pair of Army Captains from Ft. Riley ran through the introductory sequence of the program and were asked to enter initial input to the system (a set of natural language orders to subordinates in the Enemy Over the Bridge Scenario).

Evaluators were initially presented with a combination of a verbal/slide presentation and demonstrations of the running ComMentor prototype. In the later evaluation, all evaluators other than the Ft. Knox ARI staff were presented with only the running prototype. The Army Captains were provided an opportunity to work with the program directly, while one of our SMEs and one of our Klein teammates observed. In addition to taking notes on evaluator feedback, an audiotape and session logs were captured we the Army Captains used the prototype system at Ft. Riley.

**Lessons Learned**

***Does the system offer an appropriate kind of interaction, and supporting visualizations?*** This question was explored from two angles: (1) Is the mentoring being offered by our SMEs appropriate to the goal of improving battlefield command reasoning skills? (2) Is the ComMentor system appropriately designed to recreate important aspects of those live mentored sessions? From the perspective of our ARI sponsors and our many active duty officer students, it is clear that our SMEs are providing exceptionally high quality training—training that is, in general, totally unavailable to the average Captain. Our system design captures many of the aspects of those live sessions, as described earlier in this paper. The combination of natural language and map-based input and output has been identified as particularly appropriate and natural.

***Is the interface design and layout clear?*** We have been through many iterations of the interface design, most often driven by comments from our SMEs. For instance, the fact that the situation map so dominates the interface is due to their advice. The introduction of a feedback step when entering text (and, for unrecognized words that may indicate spelling errors, the use of now conventional visual cues combined with right-mouse options for alternate words) are features that seem to be clear and work well for prospective users. There were, however, a large number of requests for enhancements in this area, ranging from adding grid lines to the maps, to including more animations in the introductory scenario sequence.

***How close is the prototype to being able to handle the kinds of input students produce?*** On this point we have very limited data. Of our two Army Captains, one (as alluded to earlier) froze when presented with the "Enemy Over the Bridge" TDG. Thus no data directly bearing on this point was collected in that session. Our other Captain was very experienced, and immediately typed in a series of orders. In this case, the system correctly interpreted the first two student inputs, and missed the third only due to a known syntactic limitation of the parser. We conclude from this tiny sample (1) that, given Captains' wide range of experience, there are likely to be some that it may simply be too difficult to support (especially at the low end of spectrum), and (2) for those in our intended range, we will probably be able to interpret much of what they type at the system.

**Discussion**

Input received during these two rounds of formative review have reinforced some design decisions, and led to system modifications (or plans for modifications) in other areas. Here we review some of the more significant confirmations and modifications:

***Likely feasibility of natural language input.*** As noted earlier, on the basis of our extremely limited experience, we are encouraged that we will be able to make the system understand enough of what students type into it to enable useful tutoring. While working with the Captains at Ft. Riley we ran a series of three TDG sessions entirely in textual form. The results give us reassurance that the tutoring style we have primarily observed in verbal form can be translated to a textual form, and also provide additional data to work with in building our parser.

***Shift from Diagnostic Cases to (Possibly Clustered) Evaluations.*** Feedback, primarily from our SMEs, helped us to move from an initial design conception based on categorizing student solutions into a small set of alternate cases, to our current approach where a larger number of smaller grain-size evaluations can be applied to student solutions. This reflects the deep understanding that expert mentors can develop of students' strengths, weaknesses, understandings, and oversights. While we do not expect ComMentor to achieve such depth of understanding, we have shifted our design to make finer-grained interpretation possible when reliable recognizers can be written.

***Introduction of Explicit Temporal Visualization.*** Again based on SME feedback, we expect to introduce, in the near future, a new supporting visualization in the interface—a timeline of scenario events. Especially given the central importance of the instructional themes "Prioritize and take action in time" and "Visualize a changing situation" we expect a timeline will both make interaction with the system clearer, and help students learn to think through situations in time.

## CONCLUSION

As of June 2002, ComMentor is about eight months into a 24-month development cycle. At this point, we have completed our scheduled human expert mentor tutorial observations, and assembled partial first prototypes of both the tutor itself, and of the authoring tools needed to create the domain and scenario content.

All of our human expert mentor tutorial sessions have been recorded, and are being transcribed and analyzed. The results of this analysis will provide us with information as to the typical flow, if any, in how our mentors address the key command themes, and how the themes interrelate during tutoring. The results will be used to develop a model of how tactical expertise is developed. In addition, each session is being analyzed for the instructional strategies used by the mentors. Those strategies are being cataloged and a model of instructional strategies and their relationship to the battlefield themes is being developed. These further analyses of the tutorial sessions will be used to refine the tutorial structure of ComMentor.

Over the next year we will also work to flesh out the prototype tutor and accompanying authoring tools, and to encode sufficient content to allow the tutor to offer solid instruction on our six Phase II scenarios.

The final months of the project will be devoted to evaluation of the system's tutoring, and potentially also of the power of the accompanying authoring tools to enable tactical experts to create new scenarios, thereby lowering the costs of system growth and maintenance.

Ongoing complementary efforts include a related Phase II SBIR with a different ITS approach, and a computer-based program developed by ARI that human instructors are using in the Armor Captains Course at Fort Knox's University of Mounted Warfare. These efforts are reported on elsewhere in the conference proceedings (Shadrick & Lussier, 2002; Ryder, Graesser, McNamara, Karnavat, & Popp, 2002).

## ACKNOWLEDGEMENTS

This work was supported by the U.S. Army Research Institute, Fort Knox, Kentucky. Funding was provided under the Cognitive Readiness SBIR program by the Office of the Secretary of Defense. The authors wish to acknowledge the tutoring conducted by our collaborators, Major (ret. USMC) John Schmitt and Brigadier General (ret. USMC) Keith Holcomb whose collective expertise in training tactical thinking forms the basis for our research and development.

## REFERENCES

Deckert, J. C., Entin, E. B., Entin, E. E., MacMillan, J., & Serfaty, D. (1994). Military Decisionmaking expertise: Final Report. Fort Leavenworth, KS: Army Research Institute.

Ericsson, K. A., Krampe, R., & Tesch-Römer, C. . (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. Psychological Review, 100, (3), 363-406.

Ericsson, K. A. (1996). The Acquisition of Expert Performance. In K. A. Ericsson (Ed.), The Road to Excellence. (pp. 1-50). Mahwah, NJ: Erlbaum Associates.

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

Ong, J. & Ramanchandran, S. (2000). Intelligent Tutoring Systems: The What and the How. *ASTD Learning Circuits*. Feb, 2000. Available at http://www.learningcircuits.org/feb2000/ong.html

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

Ryder, J. M., Graesser, A. C., McNamara, J., Karnavat, A. & Popp, E. (2002). *A dialog-based intelligent tutoring system for practicing command reasoning skills.* Proceedings from Interservice/Industry Training, Simulation and Education Conference. [On-line]. http://www.iitsec.org

Schmitt, J. F. (1994). Mastering tactics . Quantico, VA: Marine Corps Association.

Schmitt, J. F. (1996). Designing good TDGs. Marine Corps Gazette.

Shadrick, Scott B., & Lussier, James W. (In Preparation). *Think Like a Commander: Captain's Edition - Prototype 1.0.* (ARI Research Product). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Shadrick, Scott B. & Lussier, James W., (2002). *Think Like a Commander at the Armor Captain's Career Course*. Proceedings from Interservice/Industry Training, Simulation and Education Conference. [On-line]. http://www.iitsec.org

U.S. Army Research Institute (2001). Think Like a Commander [Computer CD-ROM and Materials]. (Available from the U.S. Army Research Institute for the Behavioral and Social Sciences, 851 McClellan Ave, Fort Leavenworth, KS 66072).