PERGAMON

Computers in Human Behavior 18 (2002) 729–744

Computers in Human Behavior

www.elsevier.com/locate/comphumbeh

Training individuals for distributed teams: problem solving assessment for distributed mission research☆

Stephen M. Fiore*, Haydee M. Cuevas, Sandro Scielzo, Eduardo Salas

Team Performance Laboratory, Department of Psychology, Institute for Simulation and Training, and University of Central Florida, P.O. Box 161390, Orlando, FL 32816-1390, USA

Abstract

In this paper we describe an effort investigating the feasibility and utility of cognitively diagnostic assessment of problem solving when training for distributed team tasks. We utilized computer-based knowledge elicitation methods to assess both relational problem solving, requiring the semantic integration of concepts, and dynamic problem solving, requiring the ability to integrate and apply these concepts. Additionally, we addressed how metacognitive processes interact with learning outcomes when training for complex synthetic task environments. We find first, that multiple methods of assessing problem solving performance are diagnostic of knowledge acquisition for a complex synthetic team task, and second, that general metacomprehension predisposition is related to metacomprehension accuracy in synthetic task environments. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Problem solving; Knowledge structures; Mental models; Metacognition; Card sorts; Distributed teams; Synthetic task environments

1. Introduction

With rapid advances in technology and changes to industrial operations, the problems faced by organizations in both industry and the military have become ever more challenging and complicated. In order to meet these new demands, the structure of teams in these organizations has likewise increased in complexity to include

☆ The views herein are those of the authors and do not necessarily reflect those of the organizations with which the authors are affiliated.

* Corresponding author.
E-mail address: sfiore@pegasus.cc.ucf.edu (S.M. Fiore).

PII: S0747-5632(02)00027-4

both co-located and distributed team members (e.g. McNeese, Salas, & Endsley, 2001). As such, team researchers have begun focussing their efforts on developing synthetic task environments that allow them to investigate the complexities of distributed team interaction and problem solving performance within a controlled setting (Elliott, Dalrymple, Regian, & Schiflett, 2001). Along these lines, we seek to develop an automated training and problem solving performance assessment system. Problem solving performance during distributed team training can be construed as the rapid diagnosis of a dynamic situation and the application of multiple concepts from the training material. In this paper we describe an investigation of the feasibility and utility of automating cognitively diagnostic assessment of problem solving performance for distributed team tasks.

1.1. Computer-based problem solving assessment in synthetic task environments

Because of the inherent complexity of distributed synthetic task environments, a multi-faceted approach to both training and knowledge assessment is critical if training success is to be effectively interpreted. Specifically, successful problem solving performance in these environments requires, not only the basic knowledge of how to perform the various tasks, but also a higher level conceptual understanding of how this knowledge is applied in order to optimally select the appropriate strategies and actions to meet task objectives. As such, there needs to be a variety of methods of testing such that the training system is capable of diagnosing multiple levels of understanding to identify specific trainee deficiencies.

To adequately capture these knowledge requirements, we assess learning via two distinct forms of computer-based problem solving tasks. Both methods follow the view of learning as knowledge construction, where successful learning is said to involve the active integration of concepts (e.g. Baker & Mayer, 1999; Mayer, 1996), and, as such, the assessment of learning must adequately capture this integration. Education researchers argue that many assessments fail because they only tap retention of information pertaining to a given instructional environment (e.g. Glaser & Baxter, 2000). Because of the limited applicability of such methods, some suggest that what is required are methods that assess the degree to which participants can more elaborately apply knowledge (e.g. Mayer, 1996). For example, standard declarative knowledge questions, while indicative of retention, may not demonstrate how well participants can apply that knowledge. Along these lines, in a study of complex task training in aviation, Fiore, Oser, and Cuevas (in press) utilized multi media-based problem solving assessment to demonstrate that these methods can ascertain whether or not participants have integrated the concepts presented in training (see also Cuevas, Fiore, & Oser, 2001). Fiore et al. (in press) found that augmenting training via schematic diagrams facilitated the acquisition of conceptual knowledge as measured with animated aviation vignettes, but had no effect on declarative knowledge acquisition.

From a theoretical standpoint, we follow Glaser (1989) who argued that

> beginners' knowledge is spotty, consisting of isolated definitions and superficial understandings of central terms and concepts. With experience, these items

> of information become structured, are integrated with past organizations of knowledge...Thus, structuredness, coherence, and accessibility to interrelated chunks of knowledge become...objectives for instruction (p. 272)

We are interested in the manner in which knowledge structure development is influenced by the acquisition process and how measures of problem solving may be best used to diagnose trainee proficiency. The rationale is based partially upon the argument that if a "particular type of knowledge is associated with successful performance then training that incorporates instruction in this type of knowledge may result in performance improvements" (Rowe, Cooke, Hall, & Halgren, 1996, p. 44). As such, we suggest that, for distributed team training to be most effective, we must better determine how cognitively diagnostic assessment can take place in the absence of instructor intervention (e.g. Britton & Tidwell, 1995). Toward that end, we test diagnostic methods and predict that these distinct methods can accurately tap knowledge acquisition for complex team tasks.

1.1.1. Problem solving as knowledge construction

Our first measure involved a computer-based knowledge elicitation method to assess a form of relational problem solving requiring the semantic integration of concepts. As O'Neil (1999) highlights, a common theme among a variety of definitions of problem solving is the accurate interconnectivity or integration of concepts (e.g. task variables). We assess this via a standard form of mental model measurement (card sorting) and analyze participant performance in relation to an "expert model" of the conceptual relations. Thus, the accuracy of participants' conceptual integration is determined and used as a potential diagnostic aid. We used a computerized card sorting software system that allows participants to manipulate screen objects in a manner analogous to the more standard card sorting methodology.

1.1.2. Problem solving as knowledge application

For our second measure, we assessed learning via a dynamic, animated problem-solving task where successful resolution required the ability to integrate and apply concepts from the tutorial. These questions assess the degree to which participants have developed the interconnected knowledge structures necessary for complex task performance. Rather than requiring full simulation play, our questions are brief scenarios that tap key team behaviors. Specifically, we use video capture technology to create "simulation vignettes" to determine whether participants have not only acquired the basic knowledge for the simulation, but also, whether this knowledge has been integrated appropriately.

In addition to providing the benefit of standard methodologies (i.e. using multiple questions to increase reliability of measures), our diagnostic knowledge application task also provides dynamic assessment using multi-sensory presentation. Thus, rather than looking at knowledge acquisition in isolation, we assess the interconnectedness of the knowledge acquired and attempt to link that to actual performance in a complex team task. As such, our dynamic problem solving assessment requires participants to "apply knowledge of content and process to demonstrate more powerful command of an area'' (Baker & Mayer, 1999, p. 271), with accuracy determined by the action chosen.

1.1.3. Metacognition in synthetic task environments

Last, we additionally addressed how metacognitive processes interact with learning outcomes when training for complex synthetic task environments. Specifically, with many computer-based training programs, learning is oftentimes self-paced and independent of external control or monitoring from an instructor (Salas, Kosarzycki, Burke, Fiore, & Stone, in press). However, effective learner control over training requires that learners are adept at understanding their knowledge acquisition process, that is, are able to accurately monitor and evaluate their comprehension of the material (Ford, Smith, Weissbein, Gully, & Salas, 1998). Indeed, Salas et al. (in press) note that a recurring theme in distributed learning research is the criticality of learner characteristics such as metacognitive skills related to self-regulated learning. If trainees overestimate their comprehension, instruction may be terminated prematurely, leading to ineffective transfer of training and poor task performance (Osman & Hannafin, 1992). As such, we investigate how such processes are engaged while interacting within a synthetic task environment in order to develop a better understanding of metacognition in distributed learning environments.

Metacognition is a multidimensional phenomenon involving both knowledge of one's cognitions and regulation of those cognitions (Schraw, 1998). Knowledge of cognition refers to one's awareness and understanding of one's own thoughts and cognitive processes (Schraw, 1998). Regulation of cognition refers to one's ability to control and manipulate these processes (Schraw, 1998). We focus on *metacomprehension*, a principal component of metacognition involving both the ability to recognize a failure to comprehend (knowledge of cognition) and knowing *when* to engage in behaviors to repair this failure (regulation of cognition; Osman & Hannafin, 1992). Metacognitive skills, such as metacomprehension, have been shown to play an important role in self-regulated learning (e.g. Gourgey, 1998), problem solving (e.g. Mayer, 1998), and in the development of expertise (e.g. Sternberg, 1998).

In order to investigate the influence of metacognition on learners' control of their cognitive processes, researchers need reliable and valid instruments to assess learners' metacognitive processes. Additionally, such measures would allow researchers to examine the impact of these metacognitive processes on the learners' subsequent performance. This is particularly important in the context of distributed training environments given the lack of instructor intervention. Toward that goal, a number of questionnaires have been developed to assess various aspects of metacognition. For example, Moore, Zabrucky, and Commander (1997) developed the Metacomprehension Scale (MCS) to assess various components of reading comprehension abilities and strategies. The authors found evidence for convergent and discriminant validity for the MCS with other metacognitive questionnaires. Moreover, the MCS demonstrated better criterion-related validity, relative to other metacognitive measures, on comprehension performance in a reading task (Moore et al., 1997).

In this study, we assessed metacomprehension in two distinct ways in order to investigate the potentially important role of this construct when training with complex synthetic task environments. First, we administered the MCS in order to assess whether self-reports of one's predisposition towards metacomprehension behaviors may be diagnostic of metacomprehension accuracy. Second, we determined the accuracy of participant judgments of comprehension of the presented material to gauge the relation between problem solving performance and metacognitive judgments.

1.2. Summary

In Fig. 1, we have extracted and expanded upon the relevant sub-components of the CRESST model of problem solving as found in O'Neil (1999) in order to illustrate the connections between our research approach and their model. The italicized items represent the measures we developed and/or utilized to assess whether participants have adequately acquired and integrated the critical domain concepts as well as determine whether they can apply them. This methodology is part of a larger effort being conducted to investigate learning and performance in complex distributed team task environments.

First, we assess both content understanding and problem solving via a multimethod approach designed to tap differing forms of knowledge acquisition (concept recognition, declarative, and strategic questions) and knowledge application (integrative questions). Second, from the knowledge construction standpoint, we assess how the accuracy of our participants' mental models is related to successful knowledge acquisition and application. Third, we determine the degree to which metacomprehension predisposition and metacomprehension accuracy are related to knowledge acquisition and application in complex synthetic task environments. We describe these measures in more detail in the method section.

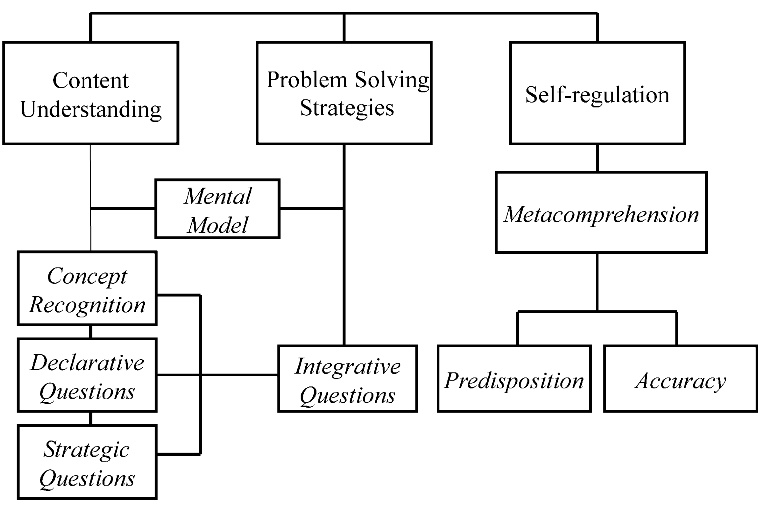

Fig. 1. Components of CRESST model of problem solving relevant to our synthetic task environment research (adapted from O'Neil, 1999).

1.3. Hypotheses

This study represents work from the development stage of our effort, and, as such, there were no specific training manipulations. Thus, we present a set of general predictions concerning, first, the diagnosticity of our problem solving measures and, second, the relation between general metacomprehension and metacomprehension accuracy in our synthetic task environment.

$\mathbf{H_1}$ (*Accuracy in participant mental model*). First, we first predict that participant accuracy in identifying conceptual relations among the concepts from the tutorial (i.e. performance in the card sorting task) will be related to performance on the knowledge acquisition measures.

$\mathbf{H_2}$ (*Accuracy on integrative problem solving assessment*). Second, we predict that participant accuracy on the dynamic problem solving task will be related to performance on the knowledge acquisition measures.

$\mathbf{H_3}$ (*Metacomprehension in synthetic task environment training*). Third, we predict that, to the degree metacomprehension behaviors generalize to learning for a complex synthetic task environment, responses on the metacomprehension scale will be related to metacomprehension accuracy with our testbed.

2. Method

2.1. Participants

Participants for this study were 25 undergraduate students (19 females and six males, mean age = 20.36) from a university in Florida. Participation in the experiment was open to all students, regardless of age, race, gender, or nation of origin. Participants received extra course credit for their participation. Treatment of these participants was in accordance with the ethical standards of the APA.

2.2. Materials

2.2.1. Distributed dynamic decision making game

Our research effort utilizes the "Distributed Dynamic Decision Making" paradigm (for a detailed description, see Kleinman & Serfaty, 1989), a synthetic task environment simulating military command and control. This game is intended to recreate the complex nature of team behavior by generating scenarios promoting team coordination and team performance (Entin, Serfaty, Elliott, & Schiflett, 2001). Within this environment, trainees own and operate various assets (e.g. helicopters, jets, tanks, and radar planes) and are individually responsible for protecting their own sector from enemy targets as well as coordinating their resources with teammates to protect the entire area. Specifically, the scenarios involve three *decision* *makers,* or team players, that have to coordinate actions (e.g. attacking targets) and share resources (e.g. transferring assets) in order to protect four sectors (i.e. their assigned sector and a shared sector) from enemy targets. Measures are based on the team's performance (e.g. overall defensive score), functioning (e.g. number of enemy targets destroyed), and processes (e.g. number of asset transfers) that are collected by the game during scenario play (Entin et al., 2001). Successful problem solving performance in this environment requires the integration of multiple knowledge formats and, thus, mimics more complex operational team tasks. For example, trainees have to master, not only declarative knowledge (e.g. range and power of their assets) and procedural knowledge (e.g. how to launch an asset or attack an enemy target), but also what we label strategic knowledge (i.e. how to apply knowledge) and integrative knowledge (i.e. the ability to use multiple task-relevant cues to interpret a situation and choose an optimal strategy).

2.2.2. Knowledge acquisition

The standard testbed training model and iterations of training for our synthetic task environment have essentially three layers: (1) introduction to domain concepts; (2) hands-on training for "how to" knowledge; and, (3) practice (free play). As part of our testbed development, we have automated the first and second layers of our testbed's training. Specifically, we developed a two-module interactive computer-based tutorial, utilizing a multimedia-format. The first module (basic declarative instruction) presented the necessary concepts for learning how to play the Distributed Dynamic Decision Making game. This included information on the key features of the game playing area, scoring, description of assets and targets, and proper rules of engagement for successful performance of the game. The second module (procedural instruction) consisted of interactive demonstrations and practice on how to perform the various tasks in the game (e.g. how to launch an asset, how to attack an enemy target, etc.). This module was designed to mimic the standard training format that typically involves an instructor watching over participants as they attempt to engage various procedures.

With our methodology, participants proceeded through this hierarchically structured tutorial at their own pace, navigating through the material using hyperlinks to access pages that provided relevant information on the concepts presented. Additionally, in the procedural training module, all participants were presented with multimedia demonstrations (i.e. audio-video-interleaved files) on how to perform the various tasks in the Distributed Dynamic Decision Making game (e.g. how to launch an asset) as well as the opportunity for interactive practice on the tasks. These interactive practice sessions are an emulated version of the domain task. This gives participants the opportunity to practice what they have seen in the multimedia presentation and provides feedback on successful and unsuccessful actions.

2.2.3. Mental model assessment—card sort task

Card sorts are a measure of knowledge structures requiring trainees to indicate how they believe concepts are related. Although a somewhat limited method, because trainees are forced to group together items rather rigidly, studies suggest that card sort data may be used to ascertain the degree to which one accurately views conceptual relations (e.g. Fiore et al., in press; Jonassen, Beissner, & Yacci, 1993). As such, we use this to assess accuracy in semantic integration of concepts from the training. Rather than follow the traditional administration of the card sort task using index cards which participants must sort manually (a task which can be laborious and time-consuming for both administration and analysis), the present study employed a fully automated card sort program. Such tools maximize the utility of computer technology for data collection and analysis by minimizing the role of the experimenter and/or instructor.

2.2.3.1. Card sort program. The TPL-KATS card sort software is a novel product developed by the Team Performance Laboratory at the University of Central Florida (Copyright 2001) to facilitate the administration of the card sort task. The user interface is composed of three main elements: *card list*, *pile list*, and *board*. The *card list* consists of the concepts defined by the administrator. The *pile list* holds the piles (i.e. named categories for the sorted cards) that the participants have created. The *board* is the workspace where the cards and piles are manipulated. The task ends when all cards have been sorted into piles and each pile has been properly labeled. For the present study, 33 concepts from our testbed tutorial were included in the card sort list. Participants were asked to group these concepts into as many piles (i.e. categories) as they thought appropriate and were then asked to label the categories that they created for each pile of cards.

2.2.4. Knowledge assessment

For the knowledge assessment task, four sets of multiple-choice questions were created that assessed concept recognition, declarative, strategic, and integrative knowledge acquisition. The first three sets were developed to match the more standard methodologies typically utilized in testing environments. The last set was designed to specifically utilize a multi-sensory and dynamic computer-based problem-solving environment that would allow us to ascertain the relation between more standard measures of the presented material and measures potentially tapping a deeper understanding of the material (cf. O'Neil, 1999). Item responses were scored simply as correct or incorrect, that is, accuracy was determined from dichotomous coding of participant responses.

1. *Concept recognition*: 10 questions ($\alpha = 0.77$) assessed participants' recognition and identification ability with respect to principle concepts from the tutorial (e.g. zones, assets, targets).
2. *Declarative knowledge*: 15 questions ($\alpha = 0.83$) assessed learners' mastery of basic factual concepts associated with the training (e.g., power and range of assets).
3. *Strategic knowledge:* 8 questions ($\alpha = 0.33$) assessed learners' ability to apply their newly acquired knowledge in various text-based scenarios. For example, participants would be presented with a description of a game scenario in

which an enemy target was entering a restricted area. The participant would have to determine, based on what they learned in the training tutorial, what would be the best action to take in this situation.

4. *Integrative knowledge:* 10 questions ($\alpha = 0.57$) used brief animated vignettes (approximately 20 s) of game scenarios, designed to tap a higher level of understanding of the material. Specifically, they assessed one's ability to integrate a variety of task-relevant cues in a testing situation that mimics the operational setting and further required that they interpret these complex and dynamic events. For example, when presented with an animated scenario where several targets were entering a restricted area, participants would have to determine what would be the best action to take, based on not only the integration of their basic declarative and procedural knowledge, but also on the status of their assets and their teammates' availability to assist.

2.2.5. *Metacomprehension questions*

We additionally solicited responses from participants so as to determine metacomprehension accuracy. Immediately following completion of the tutorial, participants were asked to rate how well they thought they would do on questions on the material presented in the tutorial. This item was rated on a scale representing performance from 0 to 100% in increments of 10 percentage points. A similar question assessed participants' metacomprehension awareness following completion of the knowledge assessment test by asking participants to report how well they thought they did on the test overall. These items were intended to measure the accuracy in self-assessment of performance for metacomprehension *prediction* and metacomprehension *postdiction*. Metacomprehension accuracy can be determined at a group level by correlating responses to this item with performance on the actual knowledge assessment task (Maki, 1998).

We additionally analyzed responses to these items by taking the absolute value of the difference between overall performance and prediction and postdiction to calculate an individualized measure of the accuracy of their self-assessments. This "bias" score was used to indicate the degree to which participants are able to monitor their understanding of complex task material, that is, the larger the bias score, the poorer one's ability to gauge understanding of the material.

2.2.6. *Metacomprehension predisposition*

The MCS, developed by Moore et al. (1997), was administered to assess the relation between general metacomprehension skills and knowledge acquisition in complex synthetic task environments. The MCS is a 22-item self-report questionnaire that assesses seven components of metacomprehension, with agreement to statements indicated on a five-point Likert-type scale (1 = *disagree strongly* to 5 = *agree strongly*). Reported reliability coefficients for the seven subscales are acceptable, ranging from 0.57 to 0.87.

2.3. Apparatus

The software programs for the Distributed Dynamic Decision Making training tutorial, knowledge assessment task, and card sort program were hosted on an IBM compatible Pentium 586 computer with a 15-inch color monitor. Participants navigated through the tutorial using a standard point-and-click mouse. The tutorial and knowledge assessment task were created on Microsoft PowerPoint 97. The TPL-KATS card sort program was written in the Java programming language. Multimedia presentation using audio-video-interleaved (AVI) files were incorporated into the procedural instruction module of the tutorial as well as within the integrative knowledge assessment questions.

2.4. Procedure

Upon arrival, participants were asked to complete an informed consent form, biographical data sheet, and various measures of individual differences (e.g. MCS). Next, all participants received computer-based instruction on the Distributed Dynamic Decision Making game using the tutorial created for this experiment and proceeded with self-paced instruction through the tutorial. Upon completion of the tutorial, the metacomprehension prediction question was administered. Participants were then asked to complete the card sort task. After the card sort task, participants were presented with the knowledge assessment task, followed by the metacomprehension postdiction questions. Finally, participants were debriefed and extra-credit was assigned. On average, the total length of the experiment was approximately 2 h.

3. Results

3.1. Mental model assessment

For the card sort task, we were interested in assessing whether the structural similarity to an expert model was diagnostic of performance on the knowledge acquisition measures in our synthetic task environment. This was administered as a relational problem solving assessment where the participants' accuracy in the semantic integration of concepts is determined. In particular, because expert knowledge consists of increased connectedness among critical concepts (e.g. Glaser, 1989), we hypothesized that participant accuracy in making such connections would be related to learning as indicated by the more standard measures of performance. For this analysis, we developed a scoring algorithm, based on signal detection theory, that allows for proficient diagnosis of critical conceptual relations. Signal detection theory posits that target identification ability (in this study, accurately detect relations among concepts) is dependent upon a combinatory process of one's *sensitivity* and *bias* towards target detection (Wickens, 1992). Sensitivity refers to the accuracy of one's detection mechanisms, in this case, the degree to which a participant is more likely to appropriately identify conceptual relations. As such, the method used in this study involves the application of a signal-detection theory analog as a means with which to analyze a participant's sensitivity to identifying conceptual relations. Specifically, by comparing data from a card sort to an expert sort, a participant's sensitivity to identifying critical conceptual relations can be determined (see Table 1). Each participant's card sort was compared with the card sort generated by our subject matter expert (i.e. the primary developer of the tutorial and knowledge assessment). The proportion of "hits" (i.e. correct conceptual relations) and false alarms (i.e. incorrect conceptual relations) were used to calculate the measure of sensitivity (i.e. d′).

In order to test the hypotheses concerning concept integration accuracy and performance, a median-split was conducted on the d′ analyses creating two groups (High_d′, $N=12$, and Low_d′, $N=12$). In support of Hypothesis 1, for the differing knowledge assessment questions, the High_d′ group significantly outperformed the Low_d′ group (refer to Table 2).

Similarly, with respect to metacomprehension accuracy as gauged by their bias scores, the High_d′ group ($M=0.13$, SD $=0.09$) had significantly lower bias scores than the Low_d′ group ($M=0.24$, SD $=0.14$), t $(22)=2.21$, $P<0.05$ (all reported t-tests are two-tailed). Taken together, these results suggest that a participant's sensitivity to critical conceptual relations in a complex synthetic task environment is significantly related to task performance as assessed by our knowledge acquisition measures.

3.2. Integrative problem solving assessment

To assess how dynamic problem solving can be used to document successful knowledge acquisition and application, a median-split was conducted on performance on the integrative questions, creating two groups (High_Int and Low_Int). Eight participants that fell on the median were dropped, leaving 17 for analysis (Low_Int = 7 and High_Int = 10). In support of Hypothesis 2, for the differing knowledge assessment questions, the High_Int group significantly outperformed the Low_Int group (refer to Table 3).

Similarly, with respect to metacomprehension accuracy as gauged by their bias scores, the High_Int group ($M=0.10$, SD $=0.12$) had significantly lower bias scores than the Low_Int group ($M=0.30$, $SD=0.09$), t $(15)=3.51$, $P<0.01$. Taken together, these results suggest that a participant's ability to integrate and apply critical

Table 1
Signal-detection analog as applied to card sort task to assess conceptual relation accuracy

		Expert decision	
		Unrelated	Related
Participant	Unrelated	*Correct rejection*	*Miss*
Decision	Related	*False alarm*	*Hit*

Table 2
Performance differences determined by mental model accuracy

Mental model similarity to expert	Question accuracy		
	Concept recognition	Declarative questions	Strategic questions
Low_d′	0.72 (0.22)	0.64 (0.22)	0.49 (0.17)
High_d′	0.95 (0.12)	0.93 (0.05)	0.67 (0.16)
	$t(22)=3.19$, $P<0.05$	$t(22)=4.23$, $P<0.05$	$t(22)=2.59$, $P<0.05$

Table 3
Performance differences determined by problem solving accuracy

Performance on dynamic problem solving	Question accuracy		
	Concept recognition	Declarative questions	Strategic questions
Low_Int	0.63 (0.20)	0.62 (0.18)	0.43 (0.14)
High_Int	0.88 (0.18)	0.85 (0.23)	0.63 (0.19)
	$t(15)=2.62$, $P<0.05$	$t(15)=2.17$, $P<0.05$	$t(15)=2.35$, $P<0.05$

concepts in the complex synthetic task environment is significantly related to task performance as assessed by our knowledge acquisition measures.

3.3. Metacomprehension

To document the relation between metacomprehension and task performance, we first correlated bias scores with actual performance. As predicted, overall metacomprehension *prediction* bias was negatively related to actual performance on the knowledge test (composite score of the knowledge assessment questions), $r=-0.77$, $P<0.01$. Metacomprehension *postdiction* bias was also negatively related to actual performance, $r=-0.45$, $P<0.05$. Thus, consistent with previous studies, participants poor in metacognitive accuracy tend to perform worse (i.e. the greater the metacomprehension bias, the lower the performance). To statistically test whether poorer metacognitive processes in synthetic task environments are indicative of poorer performance, participants' prediction bias scores were used to create two groups, high-bias ($N=12$) and low-bias ($N=12$). A t-test showed that the low-bias participants ($M=0.78$, $SD=0.04$) were significantly more accurate on the knowledge assessment than the high-bias participants ($M=0.55$, $SD=0.16$), $t(23)=4.73$, $P<0.01$.

Second, we assessed the relation between generalized metacomprehension predisposition, as measured by the MCS, and metacomprehension accuracy, as indicated by responses to the metacomprehension prediction and postdiction questions. A composite score for the MCS was computed to indicate levels of metacomprehension behaviors in our subject population. To assess the degree to which individual differences influenced metacomprehension accuracy in the synthetic task environment training, a median split was conducted, dividing participants into high metacomprehender (HiMC, $N=11$) and low metacomprehender (LoMC, $N=12$) groups. Two scores that fell on the median were dropped, leaving a total of 23 participants for this analysis. Participants' metacomprehension accuracy was determined by separately correlating their self-reported predictions and postdictions of performance with actual performance.

We hypothesized that those who report higher metacomprehension behaviors (HiMC) will be more accurate in gauging their comprehension of the knowledge in the complex task training. For the HiMC group, metacomprehension prediction was significantly correlated to actual performance on the knowledge test, $r=0.63$, $P<0.05$. However, when metacomprehension prediction was examined for the LoMC group, this correlation was not significant, $r=-0.02$, *ns*. Similarly, metacomprehension postdiction was significantly correlated to actual performance for the HiMC group, $r=0.67$, $P<0.05$, but, not significant for the LoMC group, $r=0.47$, *ns*.

Third, we investigated whether the differing populations vary in the degree to which familiarity with testing material augments metacomprehension accuracy performance (see Schwartz & Metcalfe, 1994). We examined the prediction versus postdiction bias scores separately for the high and low metacomprehension participants. For the LoMC group, the *prediction* bias score ($M=0.21$, $SD=0.14$) was significantly greater than their *postdiction* bias score ($M=0.13$, $SD=0.11$), $t\ (11)=2.32$, $P<0.05$. For the HiMC group, the difference between prediction bias ($M=0.17$, $SD=0.10$) and postdiction bias ($M=0.21$, $SD=0.04$) was not significant, $t\ (10)=1.5$, $P>0.05$.

4. Discussion

In this study we found that multiple computer-based methods of diagnosing problem solving performance can be used to assess knowledge acquisition for a complex synthetic team task. By using a form of relational problem solving which required the semantic integration of concepts, we were able to document how sensitivity to critical conceptual relations in the synthetic task environment (i.e. accurate mental models) was directly related to successful knowledge acquisition. Thus, this finding is consistent with the theme of problem solving involving accurate interconnectivity or integration of concepts (e.g. Frensch & Funke, 1995). We additionally used a dynamic, animated problem-solving task requiring the ability to integrate and apply concepts from the tutorial. As with the mental model assessment, we find that the ability to integrate and apply knowledge is diagnostic of successful knowledge acquisition. Last, by using a validated measure of metacomprehension, we found that general metacomprehension predisposition is related to metacomprehension accuracy in training for synthetic task environments. Furthermore, by determining metacomprehension bias, we document how inaccuracy in this process is directly related to poorer performance overall.

Given the criticality of rapid and coordinated problem solving in distributed team environments, trainees must acquire and be able to utilize inter-connected knowledge structures, that is, task specific and integrated long-term memory structures that can be activated and accessed during task performance. Problem solving in such environments can be considered to be the ability to identify how cues from the environment are linked to these knowledge structures and applied to resolve current situational needs. Furthermore, as shown in Fig. 1, problem solving requires multiple knowledge types (e.g. declarative) and utilizes differing cognitive processes (e.g. metacognition). As such, differing techniques to tap this knowledge must be implemented and, in this paper, we illustrate the utility of using dynamic techniques for the purposes of cognitively diagnostic assessment of problem solving in distributed team tasks.

Our findings on metacomprehension support the argument that training designers should assist trainees in their attempts to better monitor their subjective learning experience, namely their metacognitive processes (Bjork, 1994). For example, our bias scores illustrate the relation between poor metacomprehension and lower task performance. This suggests that those with higher bias may be less efficient learners in on-line environments. In particular, such learners may spend too much, or not enough, time with the material dependent upon the degree to which they over or underestimate their comprehension. To address this issue, we are currently testing how embedded augmentation may scaffold metacomprehension, that is, how instructional strategies can be embedded within the training to engage learners more actively in the knowledge acquisition process. For example, *queries* can be embedded within the training that require learners to elaborate on the presented material, prompting them to build internal associations among the concepts. As such, these learners may acquire a deeper conceptual understanding of the structural elements of the task domain, leading to better knowledge acquisition and successful problem solving performance (e.g. Ford et al., 1998; Hannafin & Land, 1997).

The findings reported here represent a portion of our efforts in developing methodologies to investigate distributed training for complex tasks. Through a systematic approach, we are able to provide a research capability to manipulate technological and task factors and assess performance when learning a complex distributed team task. With a principled use of distributed training technology, the findings from this effort have both practical and theoretical implications. First, from a practical standpoint, we document the utility of using dynamic problem solving methodologies to assess knowledge acquisition and application when training for complex distributed team tasks. These findings align with the arguments of Bennet et al. (1999) who illustrate the benefits of multimedia testing which include the broadening of "measurement of existing constructs by including content that would be difficult to test in the standard paper-and-pencil format" (p. 288). From a theoretical standpoint, we document the relationship between the accurate construction of semantic relations (i.e. mental models) and problem solving performance for complex synthetic task environments. We additionally find that metacomprehension processes involving both, one's predisposition to such processes as well as their accurate use when engaged in training for synthetic task environments, are indicative of successful knowledge acquisition.

In sum, the data gathered with our approach may be beneficial because researchers are increasingly utilizing Internet applications for distributed simulation and training (Entin et al., 2001). Both the development of our testbed, and the experimental investigation of manipulations within this testbed, are foundational to the effective implementation of such work. Specifically, by assessing the utility of distributed testing technology varying along operational relevant dimensions we may better understand the training and measurement issues critical to Internet-based synthetic task environment research for distributed teams.

Acknowledgements

This research was funded by Grant Number F49620-01-1-0214, from the Air Force Office of Scientific Research to Eduardo Salas, Stephen M. Fiore, and Clint A. Bowers.

References

Baker, E. L., & Mayer, R. E. (1999). Computer-based assessment of problem solving. *Computers in Human Behavior*, *15*, 269–282.

Bennett, R. E., Goodman, M., Hessinger, J., Kahn, H., Ligget, J., Marshall, G., & Zack, J. (1999). Using multimedia in large-scale computer-based testing programs. *Computers in Human Behavior*, *15*, 229–283.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe, & A. P. Shimamura (Eds.), *Metacognition: knowing about knowing* (pp. 185–205). Cambridge, MA: The MIT Press.

Britton, B. K., & Tidwell, P. (1995). Cognitive structure testing: a computer system for diagnosis of expert–novice differences. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 251–278). Hillsdale, NJ: LEA.

Cuevas, H. M, Fiore, S. M., & Oser, R. L. (2001). The facilitative effects of diagrams on scaffolding knowledge acquisition and metacognition in low verbal ability learners. In *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 936–940). Santa Monica, CA: HFES.

Elliott, L. R., Dalrymple, M., Regian, J. W., & Schiflett, S. G. (2001). Scaling scenarios for synthetic task environments: issues related to fidelity and validity. In *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 377–381). Santa Monica, CA: HFES.

Entin, E. B., Serfaty, D., Elliott, L. R., & Schiflett, S. G. (2001). DMT-Rnet: an Internet-based infrastructure for distributed multidisciplinary investigations of C2 performance. In *Proceedings of the 6th International Command and Control Research and Technology Symposium*. 19–21 June. Annapolis Maryland.

Fiore, S. M., Cuevas, H. M., & Oser, R. L., A picture is worth a thousand connections: the facilitative effects of diagrams on task performance and mental model development. *Computers in Human Behavior*. (in press).

Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology*, *83*, 218–233.

Frensch, P. A., Funke, J. (Eds.). (1995). *Complex problem solving. The European perspective*. Hillsdale, NJ: LEA.

Glaser, R. (1989). Expertise and learning: How do we think about instructional processes now that we have discovered knowledge structures? In D. Klahr, & K. Kotovsky (Eds.), *Complex information processing: the impact of Herbert A. Simon* (pp. 269–282). Hillsdale, NJ: LEA.

Glaser, R., & Baxter, G. P. (2000, January). *Assessing active knowledge* (Technical Report for Center for the Study of Evaluation). Los Angeles, CA: University of California.

Gourgey, A. E. (1998). Metacognition in basic skills instruction. *Instructional Science*, *26*, 81–96.

Hannafin, M. J., & Land, S. M. (1997). The foundations and assumptions of technology-enhanced student-centered learning environments. *Instructional Science*, *25*, 167–202.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: LEA.

Kleinman, D. L., & Serfaty, D. (1989, April). *Team performance assessment in distributed decisionmaking*. Paper presented at the Simulation and Training Research Symposium on Interactive Networked Simulation for Training, Orlando, FL.

Maki, R. H. (1998). Metacomprehension of text: influence of absolute confidence level on bias and accuracy. In D. Medin (Ed.), *The psychology of learning and motivation (Vol. 38)* (pp. 223–248). New York: Academic Press.

Mayer, R. E. (1996). Learners as information processors: legacies and limitations of educational psychology's second metaphor. *Educational Psychologist*, *31*, 151–162.

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, *26*, 46–63.

McNeese, M., Salas, E., Endsley, M. (Eds.). (2001). *New trends in collaborative activities: understanding system dynamics in complex environments*. Santa Monica, CA: HFES.

Moore, D., Zabrucky, K., & Commander, N. E. (1997). Metacomprehension and comprehension performance in younger and older adults. *Educational Gerontology*, *23*, 467–475.

O'Neil, H. F. (1999). Perspectives on computer-based performance assessment of problem solving. *Computers in Human Behavior*, *15*, 255–268.

Osman, M. E., & Hannafin, M. J. (1992). Metacognition research and theory: analysis and implications for instructional design. *Educational Technology Research & Development*, *40*, 83–99.

Rowe, A. L., Cooke, N. J., Hall, E. P., & Halgren, T. L. (1996). Toward an on-line knowledge assessment methodology: building on the relationship between knowing and doing. *Journal of Experimental Psychology: Applied*, *2*, 31–47.

Salas, E., Kosarzycki, M. P., Burke, C. S., Fiore, S. M., & Stone, D. L. Emerging themes in distance learning research and practice: some food for thought. *International Journal of Management Reviews* (in press).

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, *26*, 113–125.

Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe, & A. P. Shimamura (Eds.), *Metacognition: knowing about knowing* (pp. 93–113). Cambridge, MA: MIT Press.

Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: what makes an expert student? *Instructional Science*, *26*, 127–140.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York, NY: HarperCollins.