Chapter 11

Diagnosticity of Mental Models in Cognitive and Metacognitive Processes: Implications for Synthetic Task Environment Training

Sandro Scielzo
Stephen M. Fiore
Haydee M. Cuevas
Eduardo Salas
University of Central Florida
Orlando, Florida

Introduction

The last decade in team research has seen a rapid increase in the development and utilization of synthetic tasks to investigate the complexities of team interaction and performance. Synthetic task environments (STEs) are designed to mimic complex multi-role team tasks but are scaled to be distributed using more standard technology (Elliott, Dalrymple, Regian, & Schiflett, 2001). Through the utilization of STEs, researchers are able to systematically vary aspects of the task in order to train both cognitive and behavioral responses, but do so in a controlled setting. Moreover, as discussed by Kozlowski and DeShon in this volume, what is critical for training transfer is high psychological fidelity and not necessarily high physical fidelity (see also Bjork, 1994; Patrick, 1992). STEs offer the flexibility of high psychological fidelity while using cost-efficient low physical fidelity simulations. As such, with these STE systems, laboratory studies of complex task performance increase in external validity due to the increase in operational relevance (Elliott et al., 2001; Elliott, Hollenbeck, Schiflett, & Dalrymple, 2001).

Simultaneous to the proliferation of these systems is the increasing need to conduct collaborative research on distributed teams via internet-linked STEs. To meet this need, we are developing a programmatic research effort that incorporates elements of both cognitive science and educational psychology to assess how
distributed hypermedia applications can best be used in distributed training. Current versions of STEs are now available as internet-enabled collaborative task environments, which allow researchers to connect, not only players, but also trainers and observers. For example, the Distributed Dynamic Decision-Making Network (DDD-Net) is an internet-based, real-time STE that incorporates the requisite knowledge and behaviors essential to simulating complex team tasks (e.g., situation assessment, planning, sharing resources and information, coordinating actions, etc.). For more detailed information on DDD-Net see Materials Section and references in this chapter and Barnes & Elliott, 2001. DDD-Net and similar networked STEs (Barnes, Elliott, Schleff, and Stoyen, this volume) are developed to meet the research needs of modern day command and control, that is, research to understand situations where teams comprised of multiple individuals in physically distributed locations strive to coordinate actions more effectively via the use of emergent collaborative technologies’ (Miller, Price, Entin, Rubineau & Elliott, 2001, p. 390).

However, if research in distributed simulated environments is to truly transform training in complex tasks, two fundamental issues surrounding such training must be addressed. First is the potential for STEs to portray very complex and dynamic scenarios and second is the array of issues and challenges unique to distributed team research. The confluence of these factors results in increased error variance in participant performance associated with deficits in training. Our argument is that participant training and assessment, prior to actual experimentation, needs to be substantially improved if distributed simulation research is to be maximally effective. In this chapter we describe our recent efforts to address this issue and show how training for a complex scaled-world task can be transformed into an automated, hypermedia-based training and performance assessment system. This system is designed to train and test for multiple knowledge types and the integration of this knowledge utilizing a variety of dynamic computer-based methods.

Our long-term goals with this effort are to: first, understand how such systems can facilitate the acquisition and development of knowledge structures necessary for effective complex task performance; and, second, supplant traditional training such that a more uniform participant training can be developed for distributed simulated environments research (i.e., attenuate the occurrence of error that may be associated with non-automated intervention). In the context of STEs, we should maximize homogeneity of the training delivery in order to minimize knowledge variance across trainsees. We present a portion of this effort where we illustrate how an integrated, multi-disciplinary approach to research can be applied such that participants in synthetic task environments can be more effectively trained to criterion prior to moving on to the actual experimental task. We first describe the relation between mental models and complex STEs and discuss the importance of understanding metacognitive processes in such environments. Next, we present a portion of the findings from a study in order to illustrate how mental model measurement can be used in cognitively diagnostic assessment. We conclude with a discussion of the implications for training research when using STEs.
distributed training and collaborative task
by players, but also
Decision-Making
the complex team tasks
and information, n on DDD-Net see
Elliott, 2001. DDD-
it, and Stoyen, this
day command and
learns comprised of
coordinate actions
gies’ (Miller, Price,
is to truly transform
such training must
complex and dynamic
ique to distributed
cased error variance
. Our argument is
mentation, needs to
is to be maximally
ness this issue and
transformed into an
ment system. This
dge types and the
ic computer-based
how such systems
structures necessary
traditional training
ed for distributed
of error that may be
STEs, we should
imize knowledge
where we illustrate
be applied such that
fectively trained to
first describe the
the importance of
Next, we present a
our mental model
We conclude with
STEs.

Mental Models and Complex Task Training

More and more, humans have to interact with increasingly complex systems, requiring them to efficiently integrate differing types of knowledge (e.g., declarative, procedural) for successful task performance. Moreover, the development of competence and expertise with these systems is dependent upon the trainer’s ability to acquire knowledge in a highly connected and articulated way (Glaser & Baxter, 2000). Thus, in order to effectively train for complex tasks as well as to predict training outcomes, it is critical to measure how well this knowledge integration takes place in relation to complex training environments (Rouse, Cannon-Bowers, & Salas, 1992), that is, to investigate the development of trainees’ mental models.

Mental models are widespread in the literature and often associated with schemata (e.g., Brewer, 1987; Rentsch, & Hall, 1994) and knowledge structures (e.g., Dorsey, Campbell, Foster, & Miles, 1999; Koubek, Clarkston, & Calvez, 1994). Generally, mental models are defined according to the area of research for which they are applied. For example, following a cognitive orientation to defining mental models, Johnson-Laird (1983) proposed that mental models enable trainees to make inferences and predictions, to understand phenomena, to decide what action to take while controlling its execution, and to experience events by proxy. However, the key to understanding how mental models interact with complex training is first discerning how mental models develop. Glaser (1989) stated, ‘Beginners’ knowledge is spotty, consisting of isolated definitions and superficial understandings of central terms and concepts. With experience, these items of information become structured, and are integrated with past organizations of knowledge. Thus, structuredness, coherence, and accessibility to interrelated chunks of knowledge become...objectives for instruction’ (p. 272). Consistent with this view, we argue that effective training becomes a pivotal point in the acquisition of coherent and well-structured mental models in complex task environments. Accordingly, we define mental models as those underlying structures consisting of the internal connections formed among newly acquired constructs as one learns to interact with the systems for which one is being trained. As such, instructional system designers must identify the knowledge types that are closely related to training performance (Salas & Cannon-Bowers, 1997). Likewise, training principles should guide training content and processes to support accurate mental model development.

Mental Models and Knowledge Acquisition

Training systems that enable learners to build an appropriate mental model of the relations between concepts have been shown to encourage the acquisition of knowledge structures more similar to an expert model. For example, Fiore and colleagues (Cuevas, Fiore, & Oser, 2002; Fiore, Cuevas, & Oser, 2003) have explored the influence of instructional strategies in assisting trainees’ knowledge acquisition of complex systems. They consistently found that the inclusion of
diagrams in the training facilitated the development of accurate integrated knowledge consistent with mental model development. Other factors have also facilitated accurate acquisition of mental models that are indicative of performance. Research finds that certain forms of training content sequence (e.g., abstract material followed by concrete material) led to accurate mental model development when training knowledge structures for manufacturing tasks (Koubek et al., 1994). Mental models are also indicative of team performance in complex systems. For example, Rouse et al. (1992) observed that well-structured mental models generally assist team members in coordinating and adapting their actions as well as in integrating information and resources.

Another important consideration in training design is the cognitive load associated with training content and delivery, relative to initial trainee knowledge and experience. Instructional efficiency refers to the relationship between trainees' perceived cognitive load (i.e., subjective workload) and their overall task performance (Paas, Van Merriënboer, & Adam, 1994). The cognitive load each trainee experiences during knowledge acquisition may influence instructional efficiency. According to the cognitive load theory developed by Sweller and Chandler (1994), two distinct sources of load may affect learning. Intrinsic cognitive load refers to the complexity of the material itself and what degree of elaboration the material requires. For example, when many inter-related elements need to be learned, intrinsic cognitive load is high. Extrinsic cognitive load refers to the design of the delivery method. High extraneous cognitive load is present whenever the material is poorly organized and inefficiently presented. Training materials for STEs generally have a high intrinsic cognitive load in that most material needs to be efficiently integrated. Therefore, in order to maximize the extent to which trainees are able to integrate knowledge, it is crucial to limit cognitive load due to extraneous factors such as inefficient multimodal presentations (e.g., audio and visual).

We suggest that instructional design features that foster the development of accurate and well-integrated mental models during the training process may reduce cognitive load. For example, diagrams may facilitate the description and understanding of complex interrelationships. Essentially, the load on working memory and attentional systems may be reduced because structural relations are clearer when presented via diagrams, thereby increasing the efficiency of the learner's information processing (Marcus et al., 1996). For example, Cuevas et al. (2002) found that instructional efficiency was significantly enhanced when diagrams were embedded in training for a complex task (i.e., aviation). Similarly, Kalyuga, et al. (1999) found that instructional efficiency significantly improved when color-coding was added to the original material in an attempt to reduce the extraneous source of cognitive load.

**Mental Model Assessment**

In light of the aforementioned relationship between accurate mental models and knowledge acquisition (e.g., Cuevas et al., 2002; Fiore et al., 2003; Koubek et al.,
accurate integrated factors have also are indicative of task sequence (e.g., rate mental model ofing tasks (Koubek, 1994; Rouse et al., 1992), assessing mental model development as training proceeds becomes essential (Fiore, Fowlkes, Martin-Milburn, & Oser, 2000; Jonassen, Beissner, & Yacucci, 1993). In order to accurately evaluate how well training supports novices in their efforts to effectively integrate training concepts, that is, develop accurate mental models, appropriate metrics are needed to measure this type of knowledge acquisition (Bjork, 1994; Glaser, 1989). Knowledge elicitation techniques can be used to gauge mental model development and as a diagnostic tool to evaluate the manner in which knowledge structure development impacts performance.

A variety of qualitative and qualitative methods can be used to measure mental models, each possessing unique advantages and disadvantages. A complete review of the literature on knowledge elicitation and assessing knowledge structures is beyond the scope of this chapter (for reviews see, for example, Cooke, 1999; Hoffman, Shadbolt, Burton, & Klein, 1995). Our approach uses card sorting, as it has been successfully diagnostic of learning in a variety of our research efforts (Cuevas et al., 2002; Fiore et al., 2002). Further, Fiore et al. (2000) found a degree of convergence in mental model assessment using Pathfinder similarity ratings (Schvaneveldt, 1990) and card sorting techniques. Thus, although card sorts are a somewhat limited method because trainees are forced to group together items rather rigidly, previous research does document a relation between task performance and accurate mental model development measured via card sorts (Cuevas et al., 2002; Fiore et al., 2003). Thus, this technique is an effective tool in identifying the level of organization of core concepts (Jonassen et al., 1993) and is a reliable indicator of how novices evaluate concepts (Fiore et al., 2003), and experts view conceptual relations (Fiore et al., 2000).

Mental Models and Metacognition

Training systems need to not only support the acquisition of well-defined, integrated knowledge, but also the development of the necessary metacognitive skills for learning (Mayer, 1999). The degree to which trainees acquire an accurate assessment of their own learning or proficiency is arguably as important as their actual learning in the training program (Bjork, 1994). In this chapter, we focus on one particular aspect of metacognition, namely metacomprehension. Metacognition involves knowledge and regulation of one's cognitions (Schraw, 1998). In turn, metacomprehension involves the ability to identify a failure of one's comprehension and knowing when to correct this failure (Osman & Hannafin, 1992). Metacognitive skills such as metacomprehension ability have been shown to play an essential role in the development of expertise (e.g., Sternberg, 1998). Trainees who develop accurate mental models (i.e., acquire knowledge structures more similar to an expert model) are more likely to exhibit greater metacomprehension of their learning process (e.g., Smith, Ford, & Kozlowski, 1997).
Trainees' metacomprehension accuracy can be determined by evaluating the discrepancy between their subjective estimation of performance and actual performance, that is, their metacognitive bias. Trainees acquiring accurate mental models should be better able to monitor their comprehension, and thus, exhibit lower bias (indicating more accurate metacomprehension ability). Additionally, some data suggests that awareness of the nature of the test questions allows trainees to more accurately gauge their comprehension (Schwartz & Metcalfe, 1994). Specifically, familiarizing trainees with the nature of the questions for which they are asked to make self-evaluations of performance assists them in calibrating their metacomprehension. Thus, test taking may help trainees identify failures of their comprehension, leading to lower postdiction bias scores (i.e., self-evaluation following knowledge assessment) than prediction bias scores (i.e., self-evaluation prior to knowledge assessment) (cf. Schwartz & Metcalfe, 1994).

Exploring Mental Models in Synthetic Task Environments

Within the context of STE training, the appropriate diagnostic tools must be used to investigate the processes associated with acquiring the mental models necessary for task expertise. Next, we report the findings of a study investigating mental model diagnosticity as it relates to cognitive and metacognitive processes in training within a complex synthetic task environment. Our study utilizes the Distributed Dynamic Decision-Making (DDD) paradigm (for a detailed description, see Kleinman & Serfaty, 1989). The DDD is a synthetic task environment simulating military command and control (see MacMillan, Entin, Hess, & Paley, this volume for other applications). Within this STE, 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. Furthermore, this STE requires the integration of multiple knowledge formats and, thus, mimics more complex operational training 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).

In general, successful training outcomes would be expected in participants whose mental models were more similar to an expert mental model. Specifically, participants whose mental models were more similar to an expert mental model were expected to perform better on knowledge acquisition measures and experience less cognitive load, yielding higher instructional efficiency scores. In terms of metacognitive processes, participants having acquired accurate mental models were expected to exhibit lower bias scores (indicating more accurate metacomprehension ability). Last, following the theorizing of Schwartz & Metcalfe (1994), the knowledge assessment task itself was expected to assist...
participants in calibrating their metacomprehension, leading to lower bias scores. Further, we suggest that such an effect will manifest itself to a greater degree with participants who have initially acquired a poor mental model of the task constructs.

Method

Participants

Twenty-five undergraduate students (19 females and 6 males, mean age = 20.36) from a southeastern university participated in this study for course credit. Treatment of these participants was in accordance with the ethical standards of the APA.

Materials

Knowledge acquisition. Our goal was to create a self-paced tutorial capable of efficiently conveying all the necessary information to perform well in the DDD synthetic task environment. Specifically, a multimedia computer-based tutorial was developed for this study, using Microsoft PowerPoint®, which presented the critical knowledge components for the DDD task. Participants were free to move forward and backward through the tutorial in order to review the presented information and were also given the opportunity to review the material at the end of the tutorial. The tutorial was divided into two main modules, described next.

The first module presented basic declarative training (Figure 11.1), describing the basic fundamentals in the DDD game and instructing participants on how to operate and control various assets (e.g., helicopters, jets, tanks, and radar planes) in a simulated military command and control situation. These concepts represent the various knowledge components essential to developing an accurate mental representation of the DDD synthetic task. Specifically, seven general topics were covered in the first module (e.g., game screen, scoring, assets, targets, etc.), with hyperlinks to enhanced views that provided more detailed information about each lesson. Participants were also instructed on the importance of working together as a team in order to effectively protect the overall region, as well as being individually responsible for protecting their own sector. For example, participants learned how to share resources to adequately defend an area.

The second module focused on procedural training, utilizing interactive demonstrations followed by hands-on practice on how to perform the various actions in the DDD task. This module was designed to mimic traditional training methods used in DDD studies, but does so in the absence of experimenter intervention. Thus, by removing experimenter involvement, we are able to decrease the likelihood of inconsistencies in training that might occur when distributed research involving multiple locations is conducted. For this module, participants first viewed a short video clip (i.e., audio-video interleaved file)
demonstrating how to perform a specific DDD action (e.g., launch an asset) and then were given the opportunity to practice the action on their own. The hands-on practice sessions, developed using Microsoft PowerPoint, were designed to mimic the actual DDD synthetic task itself, and to provide direct feedback on the performance of each action. Feedback was ‘on-line’ and reinforced correct behavior (e.g., accessing the right menus to enable a certain function) while providing corrections to erroneous behavior (e.g., providing hints to successfully manage a certain action) (Figure 11.2). At the end of each practice session, trainees were given the option to practice the action again or move on to the next demonstration. Overall, participants were presented with seven procedural training scenarios, covering the basic actions necessary for performing the DDD task.

![DDD Game Screen - Home Base](image)

- The game playing grid is basically one geographic area broken down into three separate sectors.
  - Each of you has a color-coded home base inside your pie-shaped sectors.
  - Your base shows up in your image as a black rectangle with your station’s name inside (DM1, DM2, or DM3).
  - Above that is a platform numbered in either red, green, or purple.
  - Make sure you see where each of the bases are on the DDD Game Screen.

The three of you as a team will be in charge of the entire gridded area.

Figure 11.1 Example of a declarative training slide

*Mental model assessment – card sort task* For this task, 33 core concepts were chosen from the main sections of the DDD tutorial (e.g., components of the game screen, target attributes, DDD procedures, etc.). Trainees were presented with these concepts and were asked to sort them into as many or as few categories as they desired, based on the degree to which they believed the concepts were related. To facilitate administration and analysis, this study used the TPL-KATS-card sort software, a computer-based card sort program developed by the Team Performance Laboratory at the University of Central Florida (Copyright 2001).
Diagnosaticity of Mental Models

Correct!
Now, click on OK to destroy A1-214

Figure 11.2 Example of a procedural training slide with feedback

Metacomprehension assessment Following completion of the tutorial, participants were asked to predict how well they thought they would do on multiple-choice questions about the material presented in the tutorial. Responses were recorded on a 10-point Likert-type scale, ranging from 0 percent to 100 percent (metacomprehension prediction). A similar question assessed participants' metacomprehension awareness following completion of the knowledge assessment test by having participants report how well they thought they did on the test overall (metacomprehension postdiction). By determining the signed difference between actual performance and trainees' self-evaluation of performance, these self-assessments were used to calculate participants' prediction and postdiction bias scores, respectively. The larger the bias score, the less accurate participants were at estimating their understanding of the material presented in the training, indicating poorer metacomprehension.
Knowledge assessment. To measure how well trainees were able to integrate complex knowledge, we assessed learning via a four-section, computer-based test that tapped into increasingly more complex forms of knowledge. The first three sections (concept recognition, declarative knowledge, strategic knowledge) were developed to match the more standard methodologies typically utilized in testing environments, that is, participants’ ability to describe and explain components of their mental models (cf. Rouse & Morris, 1986). The last section (integrative knowledge) was designed to assess a participants’ ability to accurately combine these knowledge components and predict future events, another critical dimension of mental models (Rouse & Morris, 1986). The four knowledge assessment sections were:

- **Concept recognition**: Ten questions assessed participants’ ability to recognize the important objects (bases, assets, targets, etc.) and areas (zones, rings, etc.) used in the DDD task. For these questions, participants had to identify the concept highlighted in an image of the DDD task.

- **Declarative Knowledge**: Fifteen questions assessed participants’ knowledge of the basic fundamentals in the DDD task, such as scoring, identifying and attacking enemy targets, etc. For these questions, participants were presented with text-based descriptions about specific characteristics of the DDD task and were asked to select the correct concept or strategy being described.

- **Strategic knowledge**: Ten questions assessed participants’ ability to apply the basic procedures and strategies in the DDD task, such as how and when to transfer assets and attack enemy targets, etc. For these questions, participants were presented with text-based descriptions of hypothetical DDD scenarios and were asked to infer the best course of action to take in each situation.

- **Integrative knowledge**: Ten questions assessed participants’ ability to integrate their perceptual knowledge and knowledge of the basic fundamentals of the DDD task with task-relevant procedures and strategies to respond to actual scenarios. For these questions, participants were presented with captured animations of actual scenarios in the DDD task. In order to successfully answer these questions, participants had to integrate the differing types of knowledge they learned in the DDD tutorial and infer the future state of a dynamic DDD scenario (e.g., how to deal with the most threatening targets in a given situation) (Figure 11.3).

Procedure

Upon arrival, participants completed an informed consent form and a biographical data sheet. Participants then received computer-based instruction on the DDD task 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. Next, participants completed the card sort task, followed by the knowledge assessment task and the
able to integrate computer-based tests. The first three scores (knowledge) were utilized in testing in the components of action (integrative accurately combine critical dimension), assessment.

utility to recognize zones, rings, etc.) of to identify the knowledge of identifying and a were presented in DDD task and third.

ity to apply the idea and when to one, participants D scenarios and second.

ity to integrate elements of the on point to actual with captured to successfully using types of state of a ting targets in a

metacognition postdiction question. Finally, participants were debriefed and extra-credit was assigned. On average, the total length of the experiment was approximately 2 hours.

**DDD: Integrative Knowledge Assessment**

[Click on image to view the animation of the DDD scenario]

37. Assume you are DM1. What would be the BEST action(s) to take?

a. Identify all unidentified targets with AW-500 and wait to coordinate actions with the other DMs on future incoming targets.

b. Move TK-502 to attack Ggl-204. Then identify AN-352 with AW-500 and transfer the information to the other DMs.

c. Move TK-502 to attack Ggl-204 and launch another task to attack Ggl-206.

d. Attack Ggl-204 with TK-502 and move AW-500 closer to your base in order to identify future incoming targets.

**Figure 11.3 Example of an integrative question**

**Results**

**Analysis**

Unless otherwise specified, the data were analyzed using an independent samples t-test, one-tailed. An alpha level of .05 was used for all statistical analyses. In order to evaluate mental models, a quantitative measure was derived from the card sort data to determine the connectedness between concepts. A matrix of all possible pairings of the concepts was generated (N = 528). A value of 1 was assigned for all those pairings of concepts falling within the same category, and a value of 0 for the remaining pairings. This matrix was used to assess the accuracy of participants' mental models. Specifically, each participant's card sort data was correlated to the card sort data generated by our subject matter expert (SME) (i.e., the primary developer of the DDD tutorial and knowledge assessment). A median-split (Med = 0.25) was conducted on the SME correlation data creating two groups,
High_Similarity with the SME card sort ($M = 0.42$, $SD = 0.10$) and Low_Similarity with the SME card sort ($M = 0.12$, $SD = 0.11$).

**Mental Models and Knowledge Acquisition**

In support of our hypothesis that participants with mental models more similar to an expert mental model would perform better across all knowledge types, overall, results indicated that High_Similarity participants significantly outperformed Low_Similarity participants on all four sections of the knowledge assessment task (Table 11.1). Next, we calculated instructional efficiency (E) scores for High_Similarity and Low_Similarity participants by plotting the standardized scores on measures of mental effort ($R$) (i.e., subjective report of task difficulty) against the standardized scores on measures of performance ($P$) (e.g., declarative, integrative) (Paas & Van Merrienboer, 1993). The equation for instructional efficiency (adapted from Kalyuga et al., 1999) is $E = (P - R)/\text{SQRT}(2)$. The sign of $E$ is dependent on the values of $R$ and $P$. If $P > R$, then $E$ will be positive, indicating higher efficiency. If $P < R$, then $E$ will be negative, indicating lower efficiency. Baseline (or standard level of efficiency) is represented by $E = 0$.

**Table 11.1** Percent correct scores for high similarity and low similarity participants across knowledge types with means, standard deviations, and significance level

<table>
<thead>
<tr>
<th>% Correct</th>
<th>High_Similarity</th>
<th>Low_Similarity</th>
<th>$t^+$</th>
<th>sig *</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Recognition</td>
<td>.95</td>
<td>.12</td>
<td>.72</td>
<td>.22</td>
</tr>
<tr>
<td>Declarative</td>
<td>.93</td>
<td>.05</td>
<td>.64</td>
<td>.23</td>
</tr>
<tr>
<td>Strategic</td>
<td>.65</td>
<td>.13</td>
<td>.50</td>
<td>.13</td>
</tr>
<tr>
<td>Integrative</td>
<td>.49</td>
<td>.19</td>
<td>.35</td>
<td>.21</td>
</tr>
</tbody>
</table>

$^+df = 22$ for all reported $t$-statistics

* indicates one-tailed significance

Overall, results indicated that $E$-scores for High_Similarity participants were consistently positive whereas $E$-scores for Low_Similarity participants were consistently negative, thus supporting our hypothesis that participants with a mental model more similar to an expert mental model would show better instructional efficiency scores (Figure 11.4). Specifically, High_Similarity participants' $E$-scores were significantly higher than Low_Similarity participants' $E$-scores on the concept recognition, declarative, and strategic knowledge assessment sections (Table 11.2). However, although High_Similarity participants exhibited higher $E$-scores than Low_Similarity participants on the integrative knowledge assessment section, this difference was not significant.
Table 11.2 Instructional efficiency scores for high_similarity and low_similarity participants across knowledge types with means, standard deviations, and significance level

<table>
<thead>
<tr>
<th>E scores</th>
<th>High_Similarity</th>
<th>Low_Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Recognition</td>
<td>0.32</td>
<td>0.94</td>
</tr>
<tr>
<td>Declarative</td>
<td>0.37</td>
<td>0.81</td>
</tr>
<tr>
<td>Strategic</td>
<td>0.23</td>
<td>0.90</td>
</tr>
<tr>
<td>Integrative</td>
<td>0.12</td>
<td>0.93</td>
</tr>
</tbody>
</table>

* df = 22 for all reported t-statistics
* indicates one-tailed significance

Mental Models and Metacognition

Differences in bias for high_similarity and low_similarity participants As predicted, bias scores were lower for those participants who had acquired more accurate mental models (Figure 11.5). Overall, prediction bias for High_Similarity participants (M = .06; SD = .14) was significantly lower than for Low_Similarity participants (M = .22; SD = .13), t (22) = 2.89; p = .0045, indicating greater metacomprehension prediction accuracy. However, no significance difference was found on postdiction bias between High_Similarity participants (M = .03; SD = .13) and Low_Similarity participants (M = .10; SD = .19), t (22) = 1.02; p > .05.

Figure 11.4 Instructional efficiency scores for high_similarity and low_similarity participants across knowledge sections

Effect of test-taking on pre- and post-test bias scores We predicted that the knowledge assessment task would assist participants in calibrating their
metacomprehension by familiarizing them with the nature of the questions for which they were asked to make during self-evaluations of performance, particularly for those participants who had failed to acquire a mental model similar to an expert model. In support of our hypothesis, we found that postdiction bias scores were lower than prediction bias scores for both High_Similarity and Low_Similarity participants. As predicted, however, this difference was only significant for the Low_Similarity participants ($M_{pre} = .22; SD = .13; M_{post} = .10; SD = .19), t(11) = 2.18, p = .026$ (Figure 11.5). For the High_Similarity participants, postdiction bias scores ($M = .03; SD = .13$) were not significantly different from prediction bias scores ($M = .06; SD = .14), t(11) = 0.81, p > .05.

![Figure 11.5 Mean bias in percentage for high_similarity and low_similarity participants on overall test performance](image)

**Figure 11.5** Mean bias in percentage for high_similarity and low_similarity participants on overall test performance

**Discussion**

Overall, our study yielded a consistent pattern of results documenting the utility of mental model assessment in complex training for synthetic task environments. We found, from a cognitively diagnostic assessment standpoint, mental model accuracy assessed prior to testing was indicative of performance differences on a variety of knowledge assessment measures. These results also provided further support for the use of card sorts to measure mental models since this technique was able to significantly discriminate between high and low performers on increasingly more complex measures of knowledge components (Fiore et al., 2003; Rouse & Morris, 1986). As such, card sorts could potentially serve as a reliable tool for easily measuring mental model acquisition in the early stages of training.

We also found mental model accuracy to be indicative of the training program's instructional efficiency, suggesting that the instructional efficiency of complex training may be related to the accuracy of one's mental model. Specifically, results showed that an accurate mental model (i.e., as indicated by similarity with an expert model) was associated with more efficient instruction.
In this chapter, we demonstrated how an automated self-assessment research platform and assessment system could be effectively utilized to train complex models, such as those involved in training synthetic task environments. We showed that by using a combination of online and offline assessment, models could be effectively trained to perform well on complex tasks. The approach presented in this chapter demonstrates a potential methodology for the development of training programs for complex models. The results provide evidence for the diagnosis of mental models and suggest several implications for training within synthetic task environments. Specifically, the diagnostic system designed for training within these environments can help trainers determine how best to train the mental models of the trainees. The results also suggest that the greater the diagnostic feedback provided, the better the model's performance.
attempts to better monitor their subjective learning experience, namely their metacognitive processes (Bjork, 1994; see also, King, 1992).

Conclusion

At the beginning of this chapter, we argued that a set of fundamental issues must be resolved if training using STEs is to truly transform team research. First, because scaled-world tasks are, by their very nature, complex, sufficient training must be completed prior to actual interaction within such environments. Thus, some form of pre-process interaction (Fiore, Salas, & Cannon-Bowers, 2001; Fiore, Salas, Cuevas, & Bowers, 2003) is warranted in order for trainees to adequately develop an accurate understanding of the task demands and the team level interactions required to meet these demands within the STe itself. Second, because scaled-world tasks often employ distributed team research, this increases the complexity associated with the logistics of completing such research. These two factors, taken together, substantially increase the possibility of error variance associated with training, due to, for example, unreliability of training implementation and/or random irrelevancies in the experimental setting. Only when there exists a more uniform pre-process training in a controlled setting can distributed simulation research be maximally effective. In this chapter, we presented our recent efforts to address these issues and hope that our discussion motivates others to also begin the requisite empirical examination of training for synthetic task environments.

Acknowledgements

The views herein are those of the authors and do not necessarily reflect those of the organizations with which the authors are affiliated. This research was funded by Grant Number F46200-01-1-0214, from the Air Force Office of Scientific Research to Eduardo Salas, Stephen M. Fiore, and Clint A. Bowers. Address correspondence to Stephen M. Fiore, Team Performance Lab, 12424 Research Parkway, Room 406, Orlando, FL 32826 or via email at sfiores@peasus.ucf.edu.

References

Diagnosticity of Mental Models

197


