

STARS


University of Central Florida
STARS

Electronic Theses and Dissertations, 2004-2019

2004

Transforming Learning Into A Constructive Cognitive And Metacognitive

Haydee Cuevas
University of Central Florida

 Part of the [Psychology Commons](#)

Find similar works at: <https://stars.library.ucf.edu/etd>

University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Cuevas, Haydee, "Transforming Learning Into A Constructive Cognitive And Metacognitive" (2004).
Electronic Theses and Dissertations, 2004-2019. 174.

<https://stars.library.ucf.edu/etd/174>



**TRANSFORMING LEARNING INTO A CONSTRUCTIVE
COGNITIVE AND METACOGNITIVE ACTIVITY:
USE OF A GUIDED LEARNER-GENERATED INSTRUCTIONAL STRATEGY
WITHIN COMPUTER-BASED TRAINING**

by

HAYDEE MARIA CUEVAS
B.A. University of Central Florida, 1997
A. A. Miami-Dade Community College, 1995

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Arts and Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2004

Major Professor: Clint A. Bowers

© 2004 Haydee M. Cuevas

ABSTRACT

The present study explored the effectiveness of embedding a guided, learner-generated instructional strategy (query method), designed to support learners' cognitive and metacognitive processes, within the context of a computer-based complex task training environment (i.e., principles of flight in the aviation domain). Additionally, this study also examined the effect of varying the level of elaboration prompted by the queries. The queries were presented as "stop and think" exercises in an open-ended question format that asked learners to generate either simple (low level elaboration) or complex (high level elaboration) sentences from a list of key training concepts. Overall, results consistently highlighted the beneficial effect of presenting participants with low-level elaboration queries, as compared to the no-query or high-level elaboration queries.

In terms of post-training cognitive outcomes, participants presented with the low-level elaboration queries exhibited significantly more accurate knowledge organization (as indicated by greater similarity to an expert model), better acquisition of perceptual knowledge, and superior performance on integrative knowledge assessment involving the integration and application of task-relevant concepts. Consistent with previous studies, no significant differences in performance were found on basic factual knowledge assessment. Presentation of the low-level elaboration queries also significantly improved the training program's instructional efficiency, that is, greater performance was achieved with less perceived cognitive effort. Finally, participants presented with the low-level elaboration queries generated significantly more accurate sentences than participants presented with the high-level elaboration queries.

In terms of post-training metacognitive outcomes, participants presented with the low-level elaboration queries exhibited significantly greater metacomprehension accuracy (as indicated by significantly lower prediction bias scores, based on self-evaluations made following training) and more effective metacognitive self-regulation during training (as indicated by significantly greater observed levels of review effort).

Incorporating the high-level elaboration queries into the training consistently failed, with only a few exceptions, to produce significantly better post-training outcomes than the no-query or the low-level elaboration query training conditions. In addition, prompting participants to generate complex sentences (as opposed to simple sentences) did not result in an increased advantage to the training program's instructional efficiency, with these participants achieving only standard levels of performance relative to the perceived cognitive effort invested during training (i.e., baseline instructional efficiency). The increased cognitive processing of the training material associated with the high-level elaboration queries may have imposed too great a cognitive load on participants during their training, minimizing the cognitive resources available for achieving greater learning and higher levels of performance on the cognitive measures, as was evident with the low level elaboration training condition.

Although the beneficial effects of the query method were expected to be stronger for low verbal comprehension ability learners, no significant interaction between verbal comprehension ability and training condition was found. Still, the strong positive relationship found between verbal comprehension ability and several of the cognitive measures warrants further research to explore such potential aptitude-treatment interactions.

This study also hypothesized a differential impact of the query method on participants' self-reports of task-specific self-efficacy expectations for performance and metacognitive self-regulation. Contrary to predictions, after accounting for the variance attributable to these two factors at the trait level (as indicated by responses to the MSLQ pre-test), the query method did not have a significant differential effect on participants' task-specific (i.e., state-level) self-efficacy expectations of post-training performance and metacognitive self-regulation during training (as indicated by responses to the MSLQ post-test). However, given the significant positive correlation between these two factors revealed in this study, further research is warranted to tease apart the complex relationship between self-efficacy expectations of performance and metacognitive self-regulation, and more importantly, how these may impact the learning process.

The results of this study are discussed in terms of the theoretical implications for garnering a better understanding of the cognitive and metacognitive factors underlying the learning process. Practical implications for training design are presented within the context of cognitive load theory and the need for a multi-faceted approach to training evaluation.

This dissertation is dedicated to the memory of my parents, Aidee and Jose Cuevas, who through their hard work and considerable sacrifice gave me the best gift of all – a quality education – and instilled in me a passion for lifelong learning.

ACKNOWLEDGMENTS

The views herein are those of the author and do not necessarily reflect those of the organizations with which the author is affiliated. The research reported in this dissertation was partially supported by funding through Grant Number F49620-01-1-0214 from the Air Force Office of Scientific Research to Eduardo Salas, Stephen M. Fiore, and Clint A. Bowers.

This dissertation would not have been possible without the continuous guidance, support, and patience of my mentor, Dr. Stephen Fiore, and my advisor, Dr. Clint Bowers. I would also like to thank the other members of my committee, Dr. Eduardo Salas and Dr. Barbara Fritzsche, for their insightful comments and suggestions. A special thanks to my friend and colleague Sandro Scielzo for his constant moral support and his tireless efforts in helping develop the testbed for this dissertation. Finally, my most heartfelt appreciation I give to my partner, my best friend, Anthony Prieto, whose unconditional love has given me the strength to achieve even my most challenging goals.

TABLE OF CONTENTS

LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS AND SYMBOLS	xiv
INTRODUCTION.....	1
FOSTERING CONSTRUCTIVE COGNITIVE AND METACOGNITIVE ACTIVITIES. 3	
Metacognition.....	3
Query Method	4
Simple versus Complex Elaboration.....	6
Assessing Cognitive Outcomes in Complex Task Training.....	8
Knowledge Acquisition	8
Knowledge Organization	8
Instructional Efficiency.....	10
Summary.....	10
Assessing Metacognitive Outcomes in Complex Task Training.....	11
Metacomprehension Accuracy.....	11
Metacognitive Self-Regulation.....	13
Summary.....	14
Metacognition versus Self-Efficacy	14
Individual Differences in Learner Aptitudes	16
Present Study.....	18
Hypotheses.....	20
Cognitive Processes and Products Hypotheses.....	20
Hypothesis 1 – Knowledge Organization	20
Hypothesis 2 – Knowledge Acquisition	21
Hypothesis 3 – Instructional Efficiency.....	21
Hypothesis 4 – Individual Differences	22
Metacognitive Processes and Products Hypotheses	22
Hypothesis 5 – Metacomprehension Accuracy.....	22
Hypothesis 6 – Metacomprehension Calibration.....	22
Hypothesis 7 – Metacognitive Self-Regulation.....	23

Hypothesis 8 – Individual Differences	23
Hypothesis 9 – Metacognition versus Self-Efficacy.....	24
METHOD	25
Participants.....	25
Design.....	25
Materials	26
Aviation Training Tutorial (Knowledge Acquisition)	26
Airplane Parts.....	27
Flight Movements.....	27
Flight Instruments	28
Query Method	30
Module and Tutorial Surveys (Instructional Efficiency/Metacomprehension Accuracy)	31
Metacognitive Self-Regulation (Level of Review Effort)	33
Card Sort Task (Knowledge Organization)	33
Card Sort Program	34
Card Sort Tutorial	34
Knowledge Assessment Task	36
Factual Knowledge Assessment (Declarative Knowledge).....	36
Airplane Function Assessment (Integrative Knowledge).....	37
Concept Recognition Assessment (Perceptual Knowledge).....	39
Knowledge Assessment Questionnaire (Metacomprehension Postdiction Bias)	40
MSLQ and Aviation Training Effectiveness Questionnaire (Metacognition versus Self-Efficacy).....	40
Verbal Comprehension Ability	42
Time-on-Task.....	43
Apparatus	43
Procedure.....	44
RESULTS	46
Analyses	46
Check of Random Assignment.....	49
Effect of Query Method on Cognitive Outcomes	50
Knowledge Organization – Card Sort Task	52
Knowledge Assessment – Overall Performance.....	53
Knowledge Assessment – Declarative and Perceptual Knowledge.....	54
Knowledge Assessment – Integrative Knowledge	54

Instructional Efficiency.....	55
Effect of Query Method on Metacognitive Outcomes	58
Metacomprehension Accuracy – Level of Understanding.....	59
Metacomprehension Accuracy – Bias Scores.....	59
Metacomprehension Calibration.....	61
Metacognitive Self-Regulation – Level of Review Effort.....	62
Metacognition versus Self-Efficacy.....	64
Content Analysis of Participants’ Queries	65
DISCUSSION	68
Query Method – Low Level Elaboration	68
Query Method – High Level Elaboration	70
Metacognition versus Self-Efficacy	71
Individual Differences – Verbal Comprehension Ability.....	72
Limitations and Implications for Future Research	72
TRANSFORMING THE LEARNING PROCESS: THEORETICAL AND PRACTICAL IMPLICATIONS	74
Theoretical Implications.....	75
Learning as a Constructive Cognitive Activity.....	75
Elaboration, Learner Control, and Cognitive Load	76
Learning as a Constructive Metacognitive Activity	78
Practical Implications.....	78
Assessing Cognitive Outcomes.....	79
Assessing Metacognitive Outcomes	80
CONCLUSION	82
APPENDIX A: IRB COMMITTEE APPROVAL LETTER AND APPROVED STUDENT INFORMED CONSENT FORM.....	84
APPENDIX B: AVIATION TRAINING TUTORIAL SURVEY	90
APPENDIX C: KNOWLEDGE ASSESSMENT QUESTIONNAIRE	92

APPENDIX D: BIOGRAPHICAL DATA FORM	94
REFERENCES.....	96

LIST OF TABLES

Table 1. Intercorrelations of Verbal Comprehension Ability and Cognitive and Metacognitive Measures	47
Table 2. Unadjusted Means and Standard Deviations for Cognitive Measures by Training Condition.....	48
Table 3. Unadjusted Means and Standard Deviations for Metacognitive Measures by Training Condition.....	49
Table 4. Means and Standard Deviations for Background Variables by Training Condition	50
Table 5. Adjusted Means and Standard Errors for Knowledge Organization and Knowledge Assessment Measures by Training Condition.....	51
Table 6. Means and Standard Errors for Subjective Workload and Instructional Efficiency Scores by Training Condition	56
Table 7. Means and Standard Errors for Metacomprehension Bias Scores by Training Condition	60
Table 8. Means and Standard Errors for Time-On-Task and Level of Review Effort by Training Condition.....	63
Table 9. Means and Standard Errors for MSLQ Post-Test Measures by Training Condition.....	65
Table 10. Correlations for Sentence Accuracy with Cognitive and Metacognitive Measures	67

LIST OF FIGURES

Figure 1. Conceptual overview of the hypothesized effect of the query method on learners' cognitive and metacognitive processes.	19
Figure 2. Illustrative content of Module 1 (Airplane Parts) from Aviation Training Tutorial. ...	27
Figure 3. Illustrative content of Module 2 (Flight Movement) from Aviation Training Tutorial	28
Figure 4. Illustrative content of Module 3 (Flight Instruments) from Aviation Training Tutorial.	29
Figure 5. Illustrative content of query method (HLEQ) from Aviation Training Tutorial.....	31
Figure 6. Illustrative content of Factual Knowledge Assessment.....	37
Figure 7. Illustrative content of Airplane Function Assessment.....	38
Figure 8. Illustrative content of Concept Recognition Assessment.	39
Figure 9. Effect of query method on training program's instructional efficiency.	58

LIST OF ABBREVIATIONS AND SYMBOLS

ATI: aptitude-treatment interaction

E: instructional efficiency score

ECDS: embedded content-dependent strategies

HLEQ: high-level elaboration query condition

LLEQ: low-level elaboration query condition

MSLQ: Motivated Strategies for Learning Questionnaire

NQ: no-query condition

P: measure of performance used to calculate training program's instructional efficiency

R: subjective report of task difficulty used to calculate training program's instructional efficiency

INTRODUCTION

Rapid advances in technology and changes to industrial operations have created an unparalleled demand for training whereby traditional classroom learning approaches have been increasingly supplanted by distributed learning environments (Brown & Ford, 2002). Distributed instruction relies primarily on computer-based learner-controlled training programs. Whereas in traditional classroom instruction, an instructor monitors and evaluates trainees' progress, in distributed instructional approaches, learners control the pacing of the information and assess their own comprehension of the presented material before proceeding to the next lesson (Brown & Ford, 2002; Salas, Kosarzycki, Burke, Fiore, & Stone, 2002; Schmidt & Ford, 2001).

Successful learning outcomes in these technology-mediated instructional environments, therefore, are inherently dependent upon learners' possession of well-developed *metacognitive* skills, that is, how well learners are able to accurately monitor and regulate their knowledge acquisition process (Brown & Ford, 2002; Ford, Smith, Weissbein, Gully, & Salas 1998; Mayer, 1999; Osman & Hannafin, 1992; Salas et al., 2002; Schmidt & Ford, 2001). Poor comprehension monitoring may lead to premature termination of instruction, resulting in ineffective transfer of training and poor task performance (Osman & Hannafin, 1992). Therefore, it becomes even more critical that instructional systems designers and training personnel understand the cognitive and metacognitive processes involved in learning within such environments (Annett, 1989). In particular, for training to be successful, programs need to incorporate useful instructional strategies that prompt trainees to monitor their comprehension and elaborate on the material presented, that is, encourage trainees to take an active approach to learning (Bjork, 1994; Brown & Ford, 2002; Mayer, 1999).

Toward this end, this dissertation investigated the effectiveness of embedding a guided, learner-generated instructional strategy, designed to support learners' cognitive and metacognitive processes, within the context of a computer-based complex task training environment. The primary objective was to transform the learning experience into a *constructive* cognitive and metacognitive activity, where learners become "mindful architects of their own knowledge" (Glaser & Baxter, 2000, p. 1) rather than passive recipients of information. *Mindfulness* involves making a deliberate and systematic cognitive effort to engage the material during the learning process (Brown & Ford, 2002). In addition, this study also investigated the differential impact of instructional strategies on participants' self-reports of task-specific (i.e., state-level) self-efficacy expectations for performance and metacognitive self-regulation, after accounting for the variance attributable to these two factors at the trait level. Finally, because a growing body of research has shown how training interventions interact with differences in learner characteristics to produce differential results in learning (for a review, see Jonassen & Grawboski, 1993; Snow, 1997), a second objective was to further investigate how the learning outcomes yielded by this instructional strategy may be influenced by individual differences in cognitive aptitudes, specifically verbal comprehension ability. This next section describes the instructional strategy that was the focus of the present study, beginning with a discussion of the underlying mechanisms by which this training intervention may support learners' cognitive and metacognitive processes.

FOSTERING CONSTRUCTIVE COGNITIVE AND METACOGNITIVE ACTIVITIES

Metacognition

Metacognition has been defined as the awareness of one's own cognitive processes and the ability to understand, control, and manipulate these processes (Davidson, Deuser, & Sternberg, 1994, Osman & Hannafin, 1992). Metacognition, therefore, involves two distinct dimensions: *knowledge* of one's cognitions and *regulation* of these cognitions (Schraw, 1998). Knowledge of cognition refers to one's awareness and understanding of one's own thoughts and cognitive processes. Regulation of cognition refers to the behaviors one enacts to control and manipulate these processes, including learning strategies such as reviewing the material and self-testing one's knowledge. The present study focused on one particular aspect of metacognition, namely *metacomprehension*. Metacomprehension involves the "conscious processes of knowing *about* comprehending and knowing *how to* comprehend (Brown as cited in Osman & Hannafin, 1992, p. 85). Metacomprehension is not just limited to one's ability to recognize a failure to comprehend (knowledge of one's cognitions), but also to know *when* to engage in behaviors to remediate, or repair, this failure in comprehension once it has been recognized (regulation of one's cognitions) (Osman & Hannafin, 1992).

Metacognition plays an important role in communication and comprehension (both oral and written) (see Flavell, 1979), problem solving (e.g., Davidson et al., 1994; Davidson & Sternberg, 1998; Mayer, 1998), memory (e.g., Bjork, 1994; Brown, 1978), and self-regulated learning (e.g., Gourgey, 1998; Hofer, Yu, & Pintrich, 1998; Winne & Hadwin, 1998; Winne & Stockley, 1998). Metacognitive skills, such as metacomprehension, have also been shown to be critical to the development of expertise (Glaser, 1989; Osman & Hannafin, 1992; Smith, Ford, &

Koslowski, 1997; Sternberg, 1998) and to learning within complex task training environments (e.g., Fiore, Cuevas, Scielzo, & Salas, 2002; Ford et al., 1998). For example, Ford et al. (1998) examined how individual differences (mastery versus performance goal orientation), learning strategies (metacognitive activity and practice strategies), and training outcomes (knowledge, skilled performance, and self-efficacy expectations of performance) influenced transfer of learning to a more complex task. First and foremost, metacognitive activity was found to be significantly related to knowledge acquisition, skilled performance at the end of training, and self-efficacy expectations of performance. Second, these three training outcomes were positively related to performance on a novel transfer task.

Query Method

Given the ever growing body of research documenting the importance of metacognitive skills to successful training outcomes, the challenge, therefore, for instructional systems designers and training personnel, is to identify salient instructional strategies that effectively support these critical metacognitive processes. Toward this end, this dissertation investigated the utility of the *query method*, an instructional strategy that falls under the category of *embedded content-dependent strategies* (ECDS), following Osman and Hannafin's (1992) classification scheme for metacognitive training strategies. Per their definition, ECDS are "specialized, task-specific strategies applicable to particular content" (Osman & Hannafin, 1992, p. 91). By *embedding* queries within the lesson, learners are prompted to attend to and interact with the critical concepts in the presented material, increasing processing of the information and facilitating comprehension monitoring and knowledge acquisition. By making the queries *content-dependent*, emphasis can be placed on the lesson's concepts and their unique

interrelations, promoting learning of the target domain, which would be expected to result in better transfer task performance. In addition, the query method is based on a class of strategies presented in Jonassen and Grabowski's (1993) taxonomy of effective instructional strategies and tactics. In general, these strategies serve primarily to provide learners with active control of instruction (i.e., their knowledge acquisition process) by enabling learner elaboration of the material and eliciting metacognitive self-regulation activities. These instructional tactics are specifically aimed at prompting learners to monitor their comprehension, elaborate on the material presented, evaluate the meaningfulness of the information, and relate the strategies to learning objectives (Jonassen & Grabowski, 1993).

Furthermore, studies have shown that explicitly prompting self-generated elaboration of concepts (via self-explanation) leads to more accurate monitoring of comprehension, facilitates knowledge acquisition, and results in a greater understanding of the domain (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, de Leeuw, Chiu, & Lavancher, 1994; Weinstein & Mayer, 1986). For example, in a series of studies that examined several learning and study strategies, King (1992) demonstrated how a guided learner-generated questioning strategy (i.e., open-ended question stems; e.g., "*How are ____ and ____ related?*"), designed to prompt high-level elaboration of new material, facilitated knowledge acquisition and led to superior performance on objective and essay tests on the material, compared to learners not presented with such strategies. Rosenshine, Meister, and Chapman's (1996) meta-analysis on question-generation intervention studies reported similar beneficial effects for this type of learning strategy. Teaching students the cognitive strategy of generating questions for new material (described as a metacognitive or comprehension-monitoring activity) led to significant gains in comprehension,

as measured by standardized and experimenter-generated tests following the interventions (Rosenshine et al., 1996).

In a series of studies on the effects of verbalization, Berardi-Coletta, Buyer, Dominowski, and Rellinger (1995) provided further evidence for the importance of metacognitive processes in problem solving. The results of their studies demonstrated that the positive effects of verbalization of solutions on training and transfer performance on problem solving tasks was not due to the act of verbalization itself but to the metacognitive processes involved in the effort required to generate explanations for solution behaviors. Additionally, participants engaging in a process-oriented metacognitive approach consistently generated more sophisticated problem representations and developed more complex strategies. The authors proposed that the act of verbalizing solutions required students to “stop and think” about their progress, thus prompting the use of metacognitive processes.

Simple versus Complex Elaboration

The query method investigated in this dissertation involved a guided, sentence generation task requiring learner-generated elaboration of the concepts presented in the training, hypothesized to affect learners’ comprehension monitoring (i.e., metacomprehension), knowledge organization, and knowledge acquisition. Additionally, by varying the level of elaboration prompted by the queries, it would be possible to determine whether simply the presence of the queries was sufficient to prompt learners to monitor their comprehension or whether it was necessary to engage the learners in a higher level of elaboration for the instructional strategy to be most effective (cf. King, 1992). For example, Craik and Tulving (1975) found evidence for the benefits of elaboration on long-term memory by manipulating the

level of complexity of sentence frames presented to participants in a word-retention task. They found that cued recall for words accompanying *complex* sentences was twice as high as cued recall for words accompanying *simple* sentences.

To examine if the level of elaboration had a differential effect on learners' cognitive and metacognitive processes, the queries in this study were designed to prompt either low or high level elaboration of the material. These queries were presented as "stop and think" exercises in an open-ended question format that asked learners to generate sentences from a list of key concepts presented in the training. The low-level elaboration queries (LLEQ) prompted learners to generate a sentence using only *one* of the terms from this list (i.e., simple sentence). Although this query may prompt the learner to evaluate their comprehension of the concepts, the cognitive activity required to respond to this query was expected to be at a shallow level in that the focus was on understanding the concepts in isolation. In contrast, the high-level elaboration queries (HLEQ) prompted learners to generate a sentence that connected *three or more* concepts from the list that best described the relation among those concepts (i.e., complex sentence). The primary purpose of such elaboration was to prompt the learner to build internal associations between three or more concepts in the to-be-learned material and as such, potentially lead the learner to achieve a deeper, more integrative understanding of the information (Weinstein & Mayer, 1986).

Assessing Cognitive Outcomes in Complex Task Training

Knowledge Acquisition

If training effectiveness is to be accurately interpreted, a multi-faceted approach to assessment is critical. As such, several methods need to be employed to examine the impact of the query method on learners' cognitive processes. For example, post-training assessment focused exclusively on trainees' mastery of declarative knowledge may not be sensitive enough to detect learning gains and/or identify trainee deficiencies (Cuevas, Fiore, & Oser, 2002; Fiore, Cuevas, & Oser, 2003). In contrast, measures assessing *knowledge integration* focus on trainees' ability to effectively integrate differing knowledge components (e.g., declarative, perceptual) and apply this newly acquired knowledge in a variety of dynamic task-relevant scenarios (Cuevas et al., 2002; Fiore et al., 2003; Fiore et al., 2002). Such measures would potentially demonstrate whether trainees have acquired a more flexible, higher level of understanding of the material.

Knowledge Organization

It is also critical to assess the underlying organization of these knowledge components. *Knowledge organization* involves the degree to which elements of knowledge are interconnected and integrated within meaningful networks in long-term memory (Glaser, 1989; Jonassen, Beissner, & Yacci, 1993). The development of expertise has been shown to be critically dependent upon the degree to which information stored in long-term memory exhibits a high degree of structuredness, coherence, and accessibility to organized chunks of knowledge (Glaser, 1989). Therefore, a multi-faceted approach to assessing training effectiveness also needs to evaluate how well the training supports the development of the task-relevant knowledge

structures that allow novices to effectively manage the requisite higher-order processes as they acquire task expertise (Fiore, Cuevas, & Oser, 2000; Glaser, 1989; Kraiger, Ford, & Salas, 1993; Smith et al., 1997).

A variety of quantitative and qualitative methods have been used to measure the development of trainees' knowledge structures. Each technique presents unique advantages and disadvantages in assessment (for a discussion, see Evans, Jentsch, Hitt, Bowers, & Salas, 2001; Jonassen, Beissner, & Yacci, 1993), and may elicit distinct aspects of the trainee's knowledge organization (Dorsey, Campbell, Foster, & Miles, 1999). Common techniques include concept mapping (e.g., Dorsey et al., 1999), card sorting (e.g., Fiore et al., 2003), and Pathfinder analysis (Schvaneveldt, 1990). Trainees' responses are often compared to a referent (or expert) model of the domain, permitting the diagnosis of any misconceptions, that is, to determine if trainees have acquired an accurate and complete understanding of how critical task-relevant concepts are related.

The present study employed a card sort task to examine the effect of the query method on learners' organization of task-relevant concepts. In general, card sorts are a measure of knowledge organization 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 trainees accurately view conceptual relations (e.g., Fiore et al., 2003; Fiore et al., 2002; Jonassen, Beissner, & Yacci, 1993).

Instructional Efficiency

Another important consideration in evaluating training effectiveness is the relative *efficiency* of the instructional program in terms of the demands imposed by the training on learners' cognitive resources. Referred to as the training program's *instructional efficiency* (see Paas & Van Merriënboer, 1993), this measure of training effectiveness involves examining the observed relation between subjective cognitive effort and task performance in a particular learning condition. Cognitive (or mental) effort, as indicated by subjective ratings of mental workload, is the amount of resources allocated by the learner to meet the demands or cognitive load imposed by the task (Paas, Van Merriënboer, & Adam, 1994). Training programs need to incorporate useful instructional features into the design that increase the efficiency of the learner's information processing, such that fewer cognitive resources are required to achieve optimal levels of post-training task performance (Paas & Van Merriënboer, 1993). Within the context of a multi-faceted approach to evaluating training effectiveness, this dissertation, therefore, also examined the degree to which the query method may reduce the cognitive demands on learners' working memory and attentional resources associated with complex task training by better enabling them to build internal associations among the concepts presented. Thus, higher performance on a knowledge assessment task would be achieved with less cognitive effort exerted, resulting in higher instructional efficiency for the training program.

Summary

In sum, post-training performance is best assessed utilizing a multi-sensory and dynamic event-based testing environment that is designed to evaluate mastery of training objectives using increasingly more complex tasks, ranging from simple declarative knowledge assessment to

more complex tasks requiring knowledge integration and application (Cuevas, Fiore, Salas, & Bowers, 2004). Additionally, assessing both the structure and the content of knowledge are also hypothesized to be effective tools for evaluating training effectiveness (e.g., Rowe, Cooke, Hall, & Halgren, 1996). Methodologies assessing the accuracy and inter-connectivity of trainees' knowledge organization can provide insight into the underlying cognitive processes by which knowledge structure development, and subsequent learning occurs (e.g., Glaser, 1989; Smith-Jentsch, Ricci, Campbell, & Zeisig, 1997). Finally, it is also vital to determine the load imposed on learners' cognitive resources as they proceed through the training, that is, evaluate the training program's instructional efficiency. Accordingly, this dissertation adopted such a multi-faceted approach to more sensitively and diagnostically detect the impact of the query method on learners' cognitive processes.

Assessing Metacognitive Outcomes in Complex Task Training

Metacomprehension Accuracy

The goal of embedding content-free prompts (i.e., queries) into a complex task training program was to support learners' cognitive and metacognitive processes by encouraging them to "stop and think" about the information already presented before proceeding to new concepts in the training. This strategy was expected to facilitate calibration of learners' metacomprehension (a principal metacognitive process focused on comprehension monitoring) by drawing their attention to what they know and what they do not know (Schraw, 1998). Studies have shown that tasks that prompt self-testing of comprehension can enhance learning by inviting learners to search their memory for answers (see Weinstein & Mayer, 1986). Additionally, the increased

processing of the material brought about by this type of instructional strategy has also been shown to enhance calibration of learners' metacomprehension (Maki, Foley, Kajer, & Thompson, 1990). For example, Maki and her colleagues (1990) conducted two experiments to investigate whether increased processing of text during reading led to better *metacomprehension accuracy* (i.e., the relation between self-evaluations of future performance and actual performance). The results of their studies indicated that when participants did more active processing during reading, they were able to predict performance on text material with greater accuracy. These results provide support for how instructional strategies (e.g., queries) embedded into computer-based training programs may support learners' metacomprehension processes by eliciting interactive elaboration of the material.

Two distinct measures were utilized in this dissertation to assess learners' metacomprehension accuracy (for a review of various methods, see Maki, 1998; Schwartz & Metcalfe, 1994). The first measure focused on the *relative* accuracy of learners' metacomprehension assessments, that is, the degree to which self-evaluations of their understanding of the training material correlated with their post-training performance (e.g., Dunlosky, Rawson, & McDonald, 2002; Hall & Cremer, 2000). Specifically, the Pearson *r* correlation coefficient was calculated between learners' subjective assessments of their level of understanding of the concepts presented and their actual performance on a knowledge assessment task. As such, this first measure of metacomprehension accuracy examined the degree to which learners' self-reports of perceived learning varied with their actual post-training performance (Maki, 1998).

The second measure of metacomprehension accuracy served as a more individualized measure of the accuracy of learners' self-assessments. Specifically, the numerical difference

between prediction of future performance and actual performance was used to calculate a “bias” score (Maki, 1998). The larger the bias score, the poorer the learners’ ability to gauge their understanding of the training material. Furthermore, the direction of this difference (positive or negative) can be used to ascertain the nature of learners’ *confidence* (i.e., overconfidence or underconfidence, respectively) in their perceived learning (e.g., Fiore et al., 2002; Kelemen, Frost, & Weaver, 2000).

Metacognitive Self-Regulation

The previously described methods focused on learners’ awareness or *knowledge* of their cognitive processes, rather than on learners’ *regulation* of these processes. Yet, the increased comprehension monitoring elicited by the query method was also expected to induce learners to engage in self-regulatory behaviors as needed, such as increasing their level of *review effort*. Review effort refers to “the extent to which learners engage in deliberate and mindful examination of materials following an initial presentation” (Sorensen, Brown, Werner, & Huntley, 2001, p. 4). Learners’ level of review effort can be evaluated by specifically observing whether or not they go back and review the material in the training before proceeding to the next lesson. The guided sentence-generation task associated with the query method was expected to better enable learners to recognize failures in their comprehension of the material and prompt them to remedy this deficiency in initial learning with the appropriate self-regulation behavior (e.g., going back and reviewing the material). Such active regulation of their learning process would be expected to facilitate successful knowledge acquisition and transfer task performance. As such, observed level of review effort served as an additional measure of the impact of the query method on learners’ metacognitive processes.

Summary

In sum, as with evaluation of cognitive processes, to more effectively evaluate the impact of instructional strategies, such as the query method, on learners' metacognitive processes, it is critical to utilize several distinct yet related approaches, that may provide stronger convergent validity and aid in the interpretation of the results. Specifically, measures need to examine learners' metacognitive processes, both in terms of knowledge or awareness of their cognitions (e.g., metacomprehension accuracy) as well as regulation of their cognitions (e.g., level of review effort).

Metacognition versus Self-Efficacy

In addition to having the necessary cognitive and metacognitive skills, efficacious performance in any task also requires having the self-belief that one can successfully integrate and use these skills to perform the task, that is, one must also have high *self-efficacy* expectations about one's abilities (Bandura, 1986). Bandura (1986) defines perceived self-efficacy as "people's judgments of their capabilities to organize and execute courses of actions required to attain designated types of performances. . . a judgment of one's capability to accomplish a certain level of performance" (p. 391). In other words, self-efficacy is not so much concerned with the skills one has, but with judgments, or self-assessments, of what one can do with these skills (Bandura, 1986). Self-efficacy, thus, relates to a *performance* orientation, where the learner is asked to assess their ability to achieve a set level of performance, and as such, may impact motivation and goal setting. Self-efficacy expectations might be viewed as a *state* or *outcome* that arises from multiple influences including performance feedback, previous

performance history, vicarious experiences (i.e., observing the performance outcomes of others), social influence (e.g., peer appraisals), and emotional arousal (e.g., test anxiety) (Bandura, 1977).

Metacognitive skills, although related to self-efficacy, focus instead on the awareness of one's own cognitive processes and the ability to regulate these processes (Davidson et al., 1994, Osman & Hannafin, 1992). Metacognitive skills, such as metacomprehension, thus, may be considered in terms of a *learning-oriented process*, rather than a state or outcome, and involve planning, monitoring, and evaluation of one's cognitions (cf. Berardi-Coletta et al., 1995). Learners use their metacomprehension skills to monitor and regulate what should be a constructive cognitive activity in knowledge acquisition (Glaser & Baxter, 2000). Thus, in any learning environment, self-efficacy is a judgment or belief in one's potential ability to acquire or demonstrate knowledge, whereas metacomprehension describes the processes activated as one monitors his or her own knowledge acquisition. Investigating these metacomprehension processes may provide instructional systems designers and training personnel with insight as to *how* learners actually go about acquiring that knowledge, and thus, inform the design of appropriate training interventions.

The instructional strategy investigated in this study was hypothesized to support learners' metacomprehension processes in the learning task and not necessarily their self-efficacy expectations of post-training performance. Although metacognition has been shown to be significantly related to self-efficacy expectations (cf. Ford et al., 1998), the relationship between these two constructs is complex. The self-evaluation of one's cognitions that occurs in the comprehension monitoring process may influence self-efficacy expectations of one's capability to learn (Schmidt & Ford, 2001), but this effect may be dependent upon one's metacomprehension accuracy (Schunk as cited in Ford et al., 1998). Additionally, the self-

efficacy expectations one brings to the learning task may also influence the degree to which one engages in the use of effective metacomprehension processes (Hofer et al., 1998; Schmidt & Ford, 2001; Zimmerman, 1998).

The present study explored the effectiveness of an instructional strategy (i.e., query method) aimed at scaffolding learners' metacognitive processes by prompting them to engage in self-regulation of their learning (i.e., monitoring and evaluating their comprehension). The goal was to improve learners' metacomprehension accuracy (knowledge of cognition) by making them aware of *when* they need to engage in review of the training material, that is, improve their ability to sense and repair gaps in their comprehension (regulation of cognition). Though this training intervention may also influence learners' self-efficacy expectations of performance, this effect is presumed to be indirect via its impact on metacognition. Therefore, a stronger effect was predicted for the query method on participants' self-reports of task-specific (i.e., state-level) metacognitive self-regulation during training than on their self-reports of task-specific (i.e., state-level) self-efficacy expectations of post-training performance (after accounting for the variance attributable to these two factors at the trait level).

Individual Differences in Learner Aptitudes

The preceding review of the literature, although brief, provides ample evidence regarding the essential role of metacognitive processes in learning. More importantly, studies have shown that metacognitive skills can be enhanced through training for use in a variety of task domains including problem solving, reading comprehension, science, and mathematics, and across a variety of instructional settings including elementary, secondary, and post-secondary education as well as complex task training environments (e.g., Ford et al., 1998; Gourgey, 1998; Hartman,

2001a; 2001b; Maqsd, 1998; McInerney, McInerney, & Marsh, 1997; Schmidt & Ford, 2001; Volet, 1991). Yet, it is critical to also determine how these training interventions may interact with learner characteristics.

Work by Cuevas et al. (2002) has shown how instructional strategies can be effectively incorporated into training to support cognitive and metacognitive processes for different learners. Specifically, their study investigated how individual learner characteristics interact with training manipulations to influence knowledge acquisition and metacomprehension of complex systems. Consistent with prior research (e.g., Fiore et al., 2003; Mayer & Gallini, 1990; Mayer & Sims, 1994), their study found that incorporating diagrams into a complex task training program significantly facilitated the acquisition of integrative knowledge (i.e., integration and application of task-relevant knowledge), but had no significant effect on declarative knowledge acquisition (i.e., mastery of basic factual information). Diagrammatic presentation also significantly improved the instructional efficiency of the training (i.e., higher level of performance was achieved with less perceived mental effort during training). Additionally, presentation of diagrams in the training significantly increased metacomprehension accuracy (i.e., the trainee's accuracy in predicting performance on a knowledge assessment task as related to actual performance). More importantly, the effect of diagrammatic presentation on participants' cognitive and metacognitive processes was found to be strongest for those with low verbal comprehension ability. Further investigation of such potential aptitude-treatment interactions (ATI) between trainee abilities and training program design are clearly warranted (for a detailed discussion on this issue, see Jonassen & Grawboski, 1993; Snow, 1997).

Present Study

By transforming learning into a constructive cognitive and metacognitive activity, this dissertation proposed to achieve two main objectives with respect to training effectiveness: 1) to show how the increased cognitive processing associated with the query method's interactive elaboration of concepts may facilitate accurate knowledge organization and successful knowledge acquisition (e.g., Chi et al., 1989; Chi et al., 1994; Ford et al., 1998); and, 2) to demonstrate the effectiveness of the query method in assisting learners in calibrating their metacomprehension, both improving their metacomprehension accuracy (knowledge of cognitions) and prompting appropriate metacognitive self-regulatory behaviors (regulation of cognitions). Additionally, another important objective of this dissertation was to further investigate how learner characteristics may interact with instructional strategies when training for complex systems (for a conceptual overview, see Figure 1).

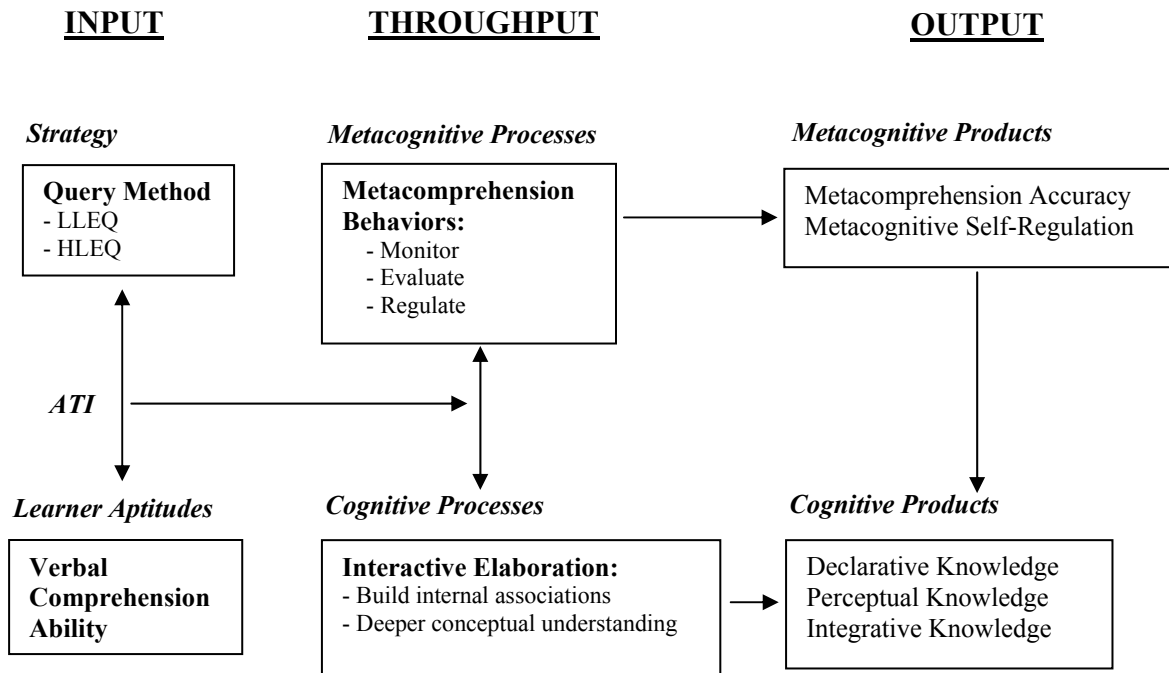


Figure 1. Conceptual overview of the hypothesized effect of the query method on learners' cognitive and metacognitive processes.

A self-paced interactive computer-based tutorial was utilized in this study, based on a complex task training testbed that required the integration of multiple knowledge formats (e.g., declarative, perceptual) and, thus, mimics complex task training. Three versions of this tutorial were developed, manipulating the presence or absence and the level of elaboration (i.e., LLEQ or HLEQ) prompted by the query method. A multi-part computer-based test was utilized to assess the influence of the query method on learners' cognitive and metacognitive processes. Specifically, three distinct forms of knowledge assessment questions measured not only learners' recognition of key concepts (perceptual knowledge) and mastery of basic factual information (declarative knowledge) associated with the training tutorial, but also the ability to integrate and

apply their newly acquired knowledge on a task involving a variety of dynamic task-relevant scenarios (integrative knowledge).

Additionally, this study also investigated the differential impact of instructional strategies on participants' self-reports of task-specific (i.e., state-level) metacognitive self-regulation and self-efficacy expectations for performance, after accounting for the variance attributable to these two factors at the trait level. Finally, because verbal comprehension ability has been found to be indicative of skill acquisition of a complex task (Fleishman & Mumford, 1989) and has also been shown to interact with learners' metacomprehension ability (e.g., Everson & Tobias, 1998; Maki, Jonas, & Kallod, 1994; Moore, Zabucky, & Commander, 1997), this dissertation also examined how the effect of the query method would be differentially influenced by individual differences in this learner aptitude.

Hypotheses

The present study examined the differential effects of the query method on cognitive and metacognitive processes and products for learners of varying levels of verbal comprehension ability. The following hypotheses were proposed:

Cognitive Processes and Products Hypotheses

This set of hypotheses pertained to the degree to which the query method influenced participants' organization and acquisition of task-relevant knowledge.

Hypothesis 1 – Knowledge Organization. Embedding queries into the training was expected to prompt participants to engage in interactive elaboration of the material. Therefore, participants presented with the query method were expected to acquire a deeper, conceptual level

understanding of the material and build more accurate internal associations among the concepts, leading to the development of well-organized knowledge structures. As such, participants presented with query method, particularly for those in the high-level elaboration (HLEQ) training condition, were hypothesized to exhibit significantly greater accuracy in their knowledge organization of the presented concepts (as measured via a card sort task) than participants not presented with the query method (*Hyp1*).

Hypothesis 2 – Knowledge Acquisition. The elaborative cognitive processing elicited by the query method was hypothesized to be more beneficial on questions assessing the integration and application of more complex types of knowledge (integrative knowledge). A lesser effect was expected on questions assessing concept recognition (perceptual knowledge) or mastery of basic factual knowledge (declarative knowledge) (cf. Cuevas et al., 2002; Fiore et al., 2003). Thus, participants presented with query method, when compared to participants not presented with this instructional strategy, were hypothesized to exhibit significantly greater performance on integrative knowledge questions (*Hyp2A*). No significant difference in performance was hypothesized for declarative or perceptual knowledge questions (*Hyp2B*).

Hypothesis 3 – Instructional Efficiency. By better enabling participants to build internal associations among the concepts presented and acquire a deeper conceptual understanding of the material, the query method was expected to improve performance on a knowledge assessment task while reducing the cognitive load associated with complex task training. As such, incorporating the query method into its instructional design was hypothesized to significantly improve the training program's instructional efficiency in relation to performance on the three sets of knowledge assessment questions (i.e., declarative, perceptual, integrative) (*Hyp3*).

Hypothesis 4 – Individual Differences. Low verbal ability learners typically have less cognitive resources available for acquisition of higher level knowledge (e.g., Davidson et al., 1994; Hartman, 2001a). However, incorporating the query method into complex task training was expected to support the development of well-organized knowledge structures and facilitate successful knowledge acquisition. As such, consistent with previous research (cf. Cuevas et al., 2002), the beneficial effects of the query method on participants’ cognitive processes was hypothesized to be stronger for those with low verbal comprehension ability (*Hyp4*).

Metacognitive Processes and Products Hypotheses

This set of hypotheses pertained to the degree to which the query method would influence participants’ comprehension monitoring (metacomprehension), both in terms of knowledge (metacomprehension accuracy) and regulation (level of review effort) of their cognitions.

Hypothesis 5 – Metacomprehension Accuracy. Embedding queries into the tutorial was expected to encourage participants to monitor and evaluate their knowledge acquisition process by requiring them to engage in active elaboration of the concepts, thereby supporting their comprehension monitoring processes (cf Maki et al., 1990). Accordingly, participants presented with the query method were hypothesized to exhibit significantly greater metacomprehension accuracy throughout their training than participants not presented with the query method, as indicated by significantly lower bias scores (*Hyp5A*) and the degree to which their perceived level of understanding varied with their actual post-training performance (*Hyp5B*).

Hypothesis 6 – Metacomprehension Calibration. Studies suggest that trainees may more accurately gauge their comprehension of targeted material if they are first made aware of the types of questions they will be asked (Schwartz & Metcalfe, 1994). In other words, familiarizing

trainees with the nature of the questions for which they are being asked to make self-evaluations of future performance may assist them in calibrating their metacomprehension (cf. Scielzo, Fiore, Cuevas, & Salas, 2004). Thus, overall, participants were hypothesized to exhibit significantly lower bias (indicating greater metacomprehension accuracy) following completion of a knowledge assessment task (*metacomprehension postdiction*) as compared to their bias scores based on self-evaluations prior to assessment (*metacomprehension prediction*), that is, postdiction bias scores were expected to be significantly smaller than prediction bias scores (*Hyp6*).

Hypothesis 7 – Metacognitive Self-Regulation. By eliciting the appropriate comprehension monitoring activities (e.g., evaluating their level of understanding of the training material), participants presented with the query method, when compared to participants not presented with these queries, were hypothesized to exhibit more effective metacognitive self-regulation behaviors (i.e., regulation of cognition), such as going back and reviewing the previous lessons to increase their understanding of the concepts. Thus, participants presented with the query method were hypothesized to exhibit significantly greater observed levels of review effort during their training than participants not presented with this instructional strategy (*Hyp7*).

Hypothesis 8 – Individual Differences. Previous studies have shown that high ability learners inherently engage in the use of metacognitive processes more than low ability learners (see Davidson et al., 1994; Osman & Hannafin, 1992; Sternberg, 1998; Weinstein & Mayer, 1986). Thus, it was hypothesized that the effect of the query method on participants' metacognitive processes would be stronger for those with low verbal comprehension ability (*Hyp8*).

Hypothesis 9 – Metacognition versus Self-Efficacy. Finally, because of the learning-oriented versus performance-oriented nature of the instructional tactics embedded in the training, it was hypothesized that the query method would have a stronger positive effect on participants' self-reports of task-specific metacognitive self-regulation during training than on their self-reports of task-specific self-efficacy expectations of post-training performance (Hyp9A). However, because a strong significant correlation has been found between metacognition and self-efficacy (e.g., Ford et al., 1998), an indirect, albeit lesser, effect was predicted for the query method on self-efficacy expectations via its impact on metacognitive self-regulation (*Hyp9B*).

METHOD

Participants

Fifty-four undergraduate students (14 males and 40 females, mean age = 21.39) from a major southeastern university participated in this experiment for course credit. Participants were recruited from the general psychology department subject pool at the university. Participation in the experiment was open to all students, regardless of age, race, gender, or nation of origin. A demographic form was used to screen out participants with previous experience with the aviation domain to ensure that only data from naïve participants were used in the analysis of the results. No participants indicated prior knowledge of the aviation domain. However, data from three participants was excluded from the analysis due to technical/procedural problems, resulting in an overall *N* of 51 (13 males and 38 females, mean age = 21.39 years). Treatment of these participants was in accordance with the ethical standards of the APA (see Appendix A for the IRB Committee approval letter). All participants received extra course credit for their participation.

Design

This study employed a one-way between-groups design, with the query method serving as the independent variable of interest. Specifically, three levels of the query method (i.e., sentence generation task) were manipulated: no query (NQ), low-level elaboration query (LLEQ), and high-level elaboration query (HLEQ). Participants were randomly assigned to conditions using a Latin squares technique to ensure that each participant had an equal chance of being assigned to any one of the three experimental groups. To investigate the impact of

individual differences on training outcomes, participants were classified post-hoc as either high or low ability based on their scores on a verbal comprehension ability measure. Multiple dependent variables were used to evaluate the effects of the query method, including knowledge organization, knowledge acquisition, instructional efficiency, metacomprehension accuracy, and metacognitive self-regulation. Additionally, a nonequivalent dependent variables design was used to assess the differential impact of the query method on participants' self-reports of task-specific metacognitive self-regulation and self-efficacy expectations of post-training performance.

Materials

Aviation Training Tutorial (Knowledge Acquisition)

To assess the manner in which the query method impacts knowledge organization and knowledge acquisition in a complex task, the present study utilized a multi-part tutorial based on the principles of flight, designed to include a variety of inter-related concepts. This tutorial was modified from the testbed created for use in our previous studies (see Cuevas et al., 2002; Fiore et al., 2003). Three versions of this interactive instructional tutorial (presented using Microsoft PowerPoint XP[®]) were developed for this experiment, with the query method manipulated at three levels: no query, low-level elaboration query, and high-level elaboration query. Material for the tutorial was adapted from the *Jeppesen Sanderson Private Pilot Manual* (1996) and the *Jeppesen Sanderson Private Pilot Maneuvers Manual* (1996), both standard training products for the instruction of pilots in the private and public sector. The tutorial was divided into three modules (Airplane Parts, Flight Movements, Flight Instruments), described next.

Airplane Parts. Module 1 described a number of airplane parts critical for standard flight operations. Participants were presented with an overview slide and 2 main slides (i.e., wings, tail), with hyperlinks to 4 additional slides that provided more detailed explanation of the concepts (e.g., ailerons, elevator) (see illustrative content in Figure 2).

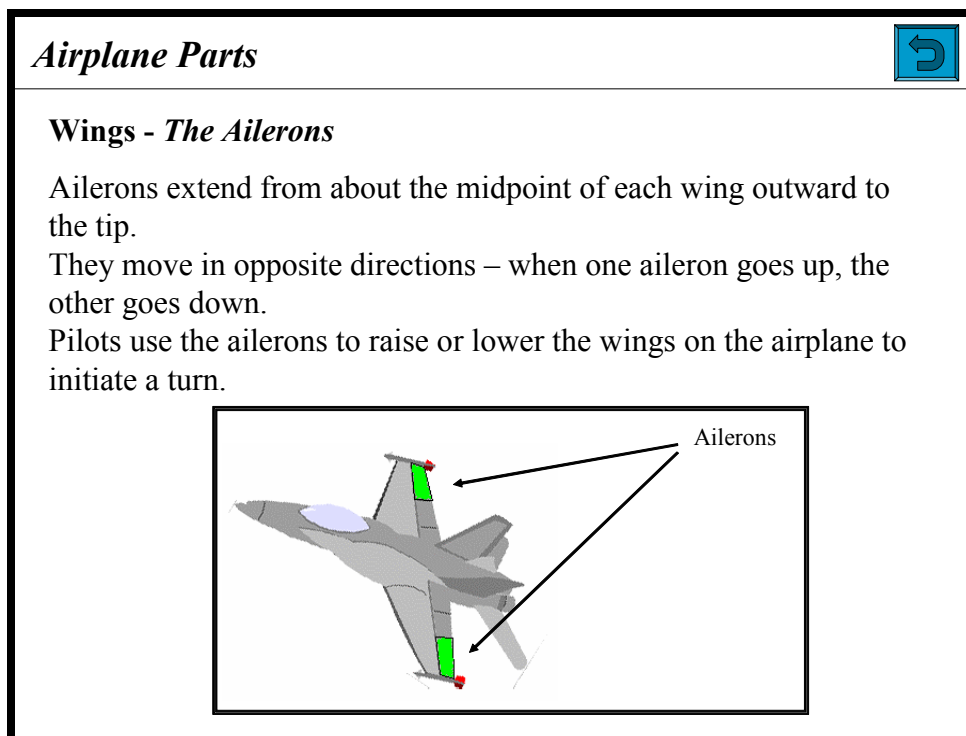


Figure 2. Illustrative content of Module 1 (Airplane Parts) from Aviation Training Tutorial.

Flight Movements. Module 2 discussed the aerodynamics of flight, including information about the axes around which an airplane moves and the movements possible in standard airplane flight. Participants were presented with an overview slide and 2 main slides (i.e., axes, movements), with hyperlinks to 6 additional slides that defined the various axes and movements (e.g., vertical axis, yaw movement) (see illustrative content in Figure 3).

Flight Movement

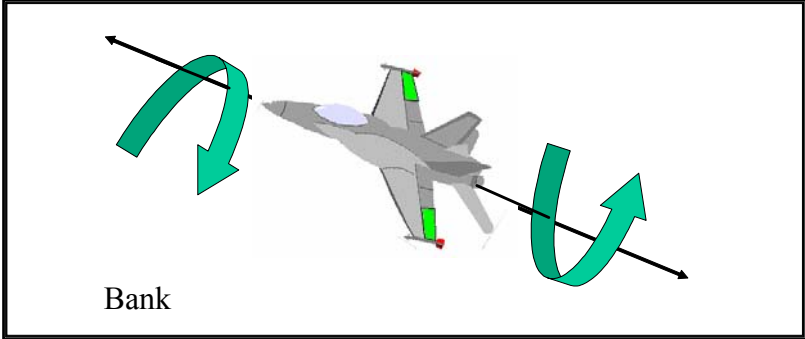
Bank Movement

Bank movement refers to the raising or lowering of the wings in reference to the horizon.

A continuous banking of the wings in a single direction is known as “roll” movement.

The pilot controls bank movement using the ailerons.


Bank movement occurs around the Longitudinal Axis.



Bank

Figure 3. Illustrative content of Module 2 (Flight Movement) from Aviation Training Tutorial

Flight Instruments. Module 3 introduced the six primary flight instruments traditionally used by pilots to navigate the airplane. Participants were presented with an overview slide and 2 main slides (i.e., pitot-static instruments, gyroscopic instruments), with hyperlinks to 12 additional slides that described how to read the instruments and explained how changes in the airplane's movements affected the information displayed on the instruments (e.g., airspeed indicator, turn coordinator) (see illustrative content in Figure 4). Additionally, participants were also presented with a hyperlink to an animated demonstration (presented using Microsoft PowerPoint XP[®]) depicting each of the six flight instruments in motion.



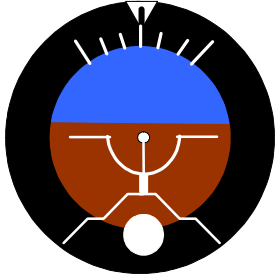
Flight Instruments

The Attitude Indicator (artificial horizon)

The Attitude Indicator (AI) is the instrument that shows the aircraft's orientation for both the nose and the wings with respect to the ground (**pitch** and **bank** attitudes respectively).

Pilots read the AI, sometimes called the "artificial horizon," when the true horizon is not visible. The AI maintains its orientation relative to the *real horizon* as the airplane banks, climbs, and descends.

The "pipper" or artificial airplane at the center of the AI is used as a point of reference. The AI has a blue "sky" and brown "earth." Because of these color indicators, when the plane climbs, the AI shows more blue and when the plane descends, the AI shows more brown.





Click here to see the AI in motion

Figure 4. Illustrative content of Module 3 (Flight Instruments) from Aviation Training Tutorial.

Participants proceeded through this hierarchically-structured tutorial at their own pace (i.e., participants were free to move backward and forward through the tutorial), navigating the hyperlinks embedded in the tutorial using a standard point-and-click mouse. No keyboard inputs were required for this portion of the experiment. All participants used hyperlinks to access pages that provided relevant information on the concepts presented. After all the lessons in the respective module had been viewed, participants in the query method conditions were presented with the guided sentence-generation task, prompting them to engage in either low or high level elaboration of the concepts (described in more detail in the next section). Participants in the no-query condition were presented with the relevant information in the lessons only.

At the end of each module, all participants were given the opportunity to go back and review the lessons before proceeding to the next module. All participants were then presented with the Module Survey (described in a later section). At the end of the tutorial, all participants were again given the opportunity to go back and review the lessons before proceeding to the next portion of the experiment. Participants were then presented with the Tutorial Survey (described in a later section). Though no time limit was imposed, participants took, on average, approximately twenty-four minutes to complete this portion of the experiment.

Query Method

Upon completion of the lessons in each module, participants in the query conditions were presented with “Stop and Think” exercises in an open-ended question format that asked them to generate sentences from a list of key concepts presented in the training. The queries were designed to prompt either low or high level elaboration of the material. The low-level elaboration queries (LLEQ) prompted participants to generate a sentence using only *one* of the terms from this list (simple sentence). The high-level elaboration queries (HLEQ) prompted participants to generate a sentence that connected *three or more* concepts from the list that best described the relation among those concepts (complex sentence) (see illustrative content in Figure 5). To demonstrate this technique, the list may include the terms: *Wings, Tail, Ailerons, Flaps, Vertical Stabilizer, Horizontal Stabilizer, Rudder, Elevator*. Choosing among these concepts, a participant in the LLEQ condition could generate the following sentence: “Pilots use the *rudder* to move the airplane’s nose left and right.” A participant in the HLEQ condition, however, could generate a sentence linking three or more of these concepts in the following

manner: “During flight, pilots initiate a turn using the *ailerons* in combination with the *rudder*, which is attached to the back of the *vertical stabilizer*.”

Stop and Think Exercises Sentence Task ◀ ▶

- Please use this exercise to increase your understanding of the material and monitor your comprehension.
- For this exercise, please look over the list of key concepts below, taken from the lessons you have just reviewed.
- Your task is to generate ONE sentence that connects *three or more* concepts from the list below that best describes the relation among those concepts.
- You are free to go back and review the material in the tutorial as you feel necessary to complete the task.
- Please write your sentence on the sheet provided in your Task Booklet then proceed to the next lesson.

<i>Wings</i>	<i>Ailerons</i>	<i>Vertical Stabilizer</i>	<i>Rudder</i>
<i>Tail</i>	<i>Flaps</i>	<i>Horizontal Stabilizer</i>	<i>Elevator</i>

Figure 5. Illustrative content of query method (HLEQ) from Aviation Training Tutorial.

Module and Tutorial Surveys (Instructional Efficiency/Metacomprehension Accuracy)

Upon completion of each module, participants were presented with a Module Survey, which asked participants to make self-assessments of the cognitive load associated with the training and predictions of performance based upon their understanding of the concepts presented in that module. Upon completion of the tutorial, participants were then presented with a Tutorial Survey, which asked similar questions as found in the Module Survey, only in this case, participants were asked to make these self-assessments of cognitive load and predictions of

performance based upon their understanding of the concepts presented in the tutorial overall (see Appendix B). As such, this last survey provided a broader assessment of participants' perceived mental workload and metacomprehension accuracy during training.

The subjective workload (cognitive load) associated with learning the training material was assessed by asking participants to report how easy or difficult they found it to understand the concepts presented in the training, with responses recorded on a 7-point scale, ranging from 1 (*very easy*) to 7 (*very difficult*). These responses were used to calculate the training's instructional efficiency (described later in the analysis of the results).

Two distinct yet related questions were used to determine participants' metacomprehension accuracy. First, participants were asked to assess their *level of understanding* of the material presented in the training, with responses, recorded on a 7-point scale, ranging from 1 (*very poor*) to 7 (*very good*). Responses to this question were correlated with participants' actual performance on the knowledge assessment task, and served as the first measure of metacomprehension accuracy. Second, participants were asked to predict, on a 10-point scale, how well they would perform on multiple-choice questions about the concepts presented in the training, with responses given in terms of percent correct, ranging from zero to one hundred percent, in ten percent increments. Responses to this question were used to calculate participants' *prediction bias scores* (i.e., the numerical difference between prediction of performance and actual performance), and served as the second measure of metacomprehension accuracy. The rationale for the use of these two distinct yet related measures was to attempt to isolate the influence of participants' metacognitive awareness, which may have a greater effect on their self-reported level of understanding, from any potential effect of self-efficacy expectations, which might influence their predictions of performance.

Metacognitive Self-Regulation (Level of Review Effort)

An experimenter observed participants as they proceeded through the tutorial to record their level of review effort, as a general indicator of their metacognitive self-regulation. Specifically, the experimenter observed whether or not the participant went back and reviewed the material before proceeding to the next lesson. Participants were assigned one of four possible levels of review effort for each module: no review (0 points), review within the module only (1 point), review upon completion of the module only (2 points), and review both within and upon completion of the module (3 points). Also noted was whether participants reviewed the lessons upon completion of the Aviation Training Tutorial (no review = 0 points; review = 1 point). The total sum of these scores was then used to calculate an overall measure of observed level of review effort during the training.

Card Sort Task (Knowledge Organization)

A card sort task was used to evaluate participants' knowledge organization of task-relevant concepts. Rather than administer 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 (described next) to maximize the utility of computer technology for data collection and analysis.

For the present study, 26 key concepts were extracted from the Aviation Training Tutorial. Participants were instructed to group these concepts into as many categories as they desired and were then asked to name or describe the categories that they created for each group of cards. Though no time limit was imposed, participants took, on average, approximately

twenty minutes to complete this portion of the experiment, including training on the software program and actual performance of the card sort task.

Card Sort Program. The TPL-KATS-card sort software is a knowledge structure assessment product developed by the Team Performance Laboratory at the University of Central Florida (Copyright 2001). Two modes of operation are available for the software, *user* and *administrator* mode. Participants perform the card sort task under the user mode, using a standard point-and-click mouse. The administrator mode is primarily for the creation of concepts that will be used as the template in the user mode to complete 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. This list is displayed in a window containing a scroll bar that allows the participant to move through the list of concepts. Participants are free to either sort the concepts into different piles (i.e., categories) and then name the piles, or name the piles first and then sort the various concepts. The *pile list* holds the piles (i.e., named categories for the sorted cards) that the participants have created. As with the card sort list, a scroll bar allows participants to review all the piles in the pile list. The *board* is the workspace where the participant manipulates the cards and piles. The task ends when all cards have been sorted into piles and each pile has been properly labeled. Though no time limit was imposed, participants took, on average, approximately ten minutes to complete card sort task.

Card Sort Tutorial. A computer-based tutorial (presented using Microsoft PowerPoint XP[®]) was utilized to train participants on how to use the card sort program. Employing a hypermedia format, the tutorial explained the basic components and procedures required to perform the card sort task. Specifically, participants were presented with both textual and

diagrammatic information on the actual card sort user interface. The displays were augmented with embedded arrows indicating movements or actions to be performed as described by the textual information. For example, if the textual information indicates that a concept has to be moved to a specific location, the diagram representing the card sort screen will show an arrow starting from the concept and pointing to the target, as indicated by the text. Information on how to perform the card sort task was presented following a sequential, progressive format (i.e., basic definitions of elements were presented first before more complex procedural information).

The card sort tutorial was divided into four main sections. The first section introduced the various components displayed on the card sort task screen (e.g., tools, list boxes, and main board). The second section described the different tools (e.g., card button, pile button, delete button) and how to properly operate them. The third section introduced and defined the concept of “piles.” In this section of the tutorial, participants learned how to move piles onto the board, how to remove them from the board, how to create temporary piles, how to name piles, how to add concepts into piles, and how to remove concepts from piles. The last section of the card sort tutorial walked the participant through a card sort example and highlighted the requisites for completing the task (e.g., sorting all concepts into named piles). Participants were able to navigate through the tutorial at their own pace, using hyperlinks that allowed them to move forward and backward through the slides and review the material as needed. Though no time limit was imposed, participants took, on average, approximately ten minutes to complete the card sort tutorial.

Knowledge Assessment Task

This study utilized three distinct forms of knowledge assessment questions in order to separately examine the influence of the query method on the acquisition of varying types of knowledge. As with the tutorial, the knowledge assessment task developed for this study was modified from the testbed created for use in our previous studies (see Cuevas et al., 2002; Fiore et al., 2003). Participants proceeded through the 48-item computer-based Knowledge Assessment Task (overall $\alpha = .859$), presented using Microsoft PowerPoint XP[®], at their own pace, proceeding from one question to the next using a standard point-and-click mouse. Factual Knowledge Assessment (i.e., declarative knowledge questions) was presented first, then Airplane Function Assessment (i.e., integrative knowledge questions), and finally, Concept Recognition Assessment (i.e., perceptual knowledge questions) (described next in detail). Only one question was presented at a time on the screen and a multiple-choice format was used for all questions. Unlike the tutorial, participants were not able to go back and review or change their responses once they had moved on to the next question, and no feedback was provided as to the accuracy of their responses. Participants, on average, completed this task in about twenty minutes.

Factual Knowledge Assessment (Declarative Knowledge). Twenty questions ($\alpha = .741$), adopted from a standard introductory flight manual (*Jeppesen Sanderson Private Pilot Exercise Book*, 1996), assessed participants' mastery of basic factual information associated with the training tutorial (e.g., definitions of the various parts of the plane). Standardized testing procedures have long relied on such assessment based upon one's effective mastery of task-relevant knowledge, and this represents a common method used in computer-based distance learning environments (e.g., Proctor & Dutta, 1995; van Oostendorp & Goldman, 1999). For this

declarative knowledge assessment task, participants were presented with text-based definitions taken from the Aviation Training Tutorial and were asked to identify the concept being described (see illustrative content in Figure 6).

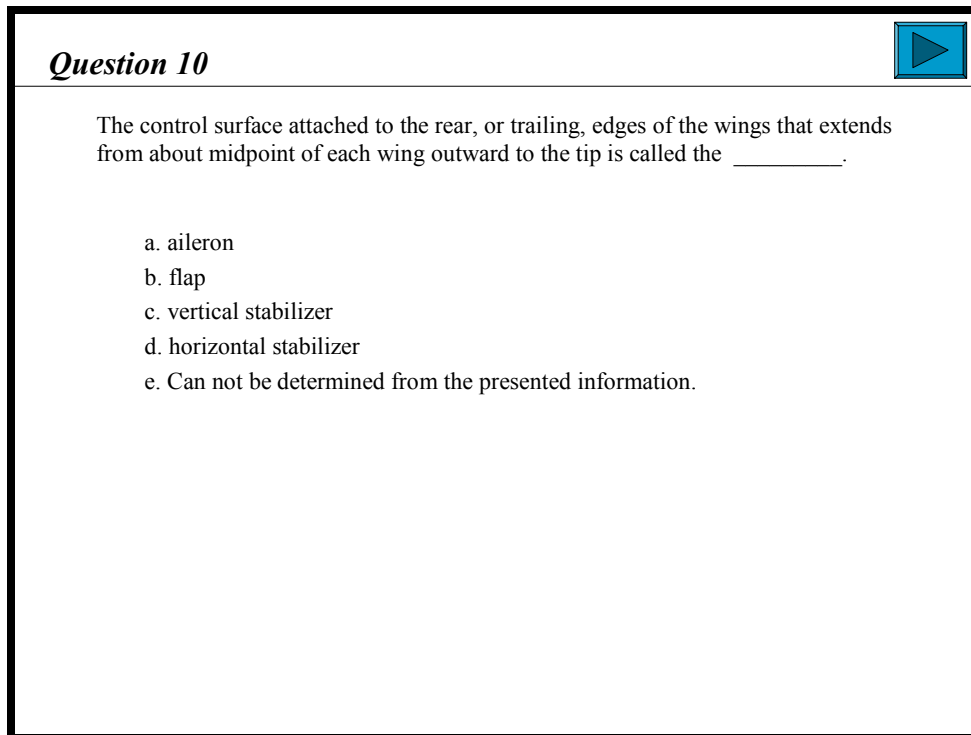



Figure 6. Illustrative content of Factual Knowledge Assessment.

Airplane Function Assessment (Integrative Knowledge). Ten questions ($\alpha = .634$) assessed participants' ability to integrate and apply task knowledge. Capitalizing on the multimedia capabilities of computer-based training and assessment systems, knowledge assessment can be designed utilizing the presentation of dynamic scenarios, rather than relying on text-based vignettes (for a discussion of "scenario-based" training, see Oser, Cannon-Bowers, Salas, & Dwyer, 1999). These questions presented participants with a variety of dynamic

animated scenarios, illustrating an application of task-relevant concepts, and required them to identify the concepts being applied. Thus, this cognitively diagnostic task measured participants' ability to integrate and apply their knowledge, rather than simply their ability to retrieve factual knowledge, such as definitions, as in the first set of questions (for a discussion on cognitively diagnostic assessment, see Fiore et al., 2002). For this integrative knowledge assessment task, participants were presented with an animated image (using audio-video interleaved file format) of an airplane performing a maneuver and were asked to determine, for example, which airplane parts and flight instruments were being utilized in the maneuver demonstrated (see illustrative content in Figure 7).

Question 28 

The pilot is using the _____ to control the airplane in this animation, which is performing a _____ movement, as indicated on the _____.

- a. ailerons; bank; attitude indicator
- b. flaps; bank; attitude indicator;
- c. ailerons; yaw; heading indicator
- d. rudder; yaw; heading indicator
- e. Can not be determined from the presented information

Click inside the window below to see the maneuver

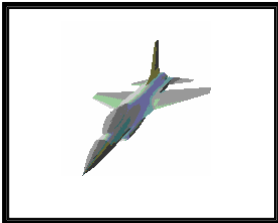



Figure 7. Illustrative content of Airplane Function Assessment.

Concept Recognition Assessment (Perceptual Knowledge). Eighteen questions ($\alpha = .675$) tested participants' perceptual knowledge with regard to their recognition ability of key concepts presented in the Aviation Training Tutorial. Participants were presented with static illustrations of the principle concepts (i.e., airplane parts, axes, movements, and instruments) from the tutorial and were asked to identify the concept depicted (see illustrative content in Figure 8). This third section, as with the declarative knowledge assessment, represents another standard form of assessment in computer-based training environments.

Question 41 

Name the highlighted airplane part.

- a. flaps
- b. elevators
- c. rudders
- d. ailerons
- e. Can not be determined from the presented information.

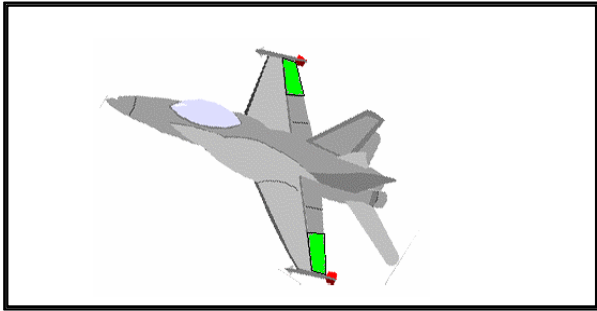


Figure 8. Illustrative content of Concept Recognition Assessment.

Knowledge Assessment Questionnaire (Metacomprehension Postdiction Bias)

Immediately following completion of the Knowledge Assessment Task, participants were asked to complete the Knowledge Assessment Questionnaire (see Appendix C). Responses to this questionnaire, similar to the Module and Tutorial Surveys described earlier, were used as an additional measure of participants' metacomprehension accuracy (cf. Fiore et al., 2002). Specifically, participants were asked to rate how well they thought they did on the Knowledge Assessment Task overall, as well as on each of the three separate sets of questions (i.e., Factual Knowledge Assessment, Airplane Function Assessment, Concept Recognition Assessment). As with the metacomprehension prediction question found in the Module and Tutorial Surveys, responses were recorded on a 10-point scale, in terms of percent correct, ranging from zero to one hundred percent, in ten percent increments. A *postdiction bias score* was calculated by taking the numerical difference between participants' postdiction of performance and their actual performance on the knowledge assessment questions.

MSLQ and Aviation Training Effectiveness Questionnaire (Metacognition versus Self-Efficacy)

Because the query method was hypothesized to impact participants' metacognitive processes during training, and not necessarily their self-efficacy expectations of performance, an independent measure of these factors was administered. The Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich, Smith, Garcia, & McKeachie, 1993) is a self-report Likert-type instrument consisting of both motivation and learning strategies scales. The motivation scales include a) three value scales: intrinsic goal orientation, extrinsic goal orientation, and task value; b) two expectancy scales: control of learning beliefs and self-efficacy for learning and performance; and c) one affect scale: test anxiety. The learning strategies scales include a) four

cognitive scales: rehearsal, elaboration, organization, and critical thinking; b) one general metacognitive scale that covers planning, monitoring, and regulating strategies; and c) four resource management scales: managing time and study environment, effort management, peer learning, and help-seeking.

For the purposes of this study, only two of the above-referenced scales were used: *Self-Efficacy for Learning and Performance* (a motivation scale) and *Metacognitive Self-Regulation* (a learning strategy scale). Reported reliability coefficients for the self-efficacy and metacognition scales are acceptable ($\alpha = .93$ and $.79$, respectively) (Pintrich et al., 1993). A significant positive correlation between the two scales has also been reported ($r = .46$) (Pintrich et al., 1993). Thus, this instrument was used to assess the potential impact of the query method on participants' self-reports of task-specific metacognitive self-regulation and self-efficacy expectations of post-training performance.

Specifically, two modified versions of the MSLQ were administered. The first modified version assessed participants' general predisposition, at the *trait* level, on these factors and was administered as a pre-test before the training to establish baseline data. The second modified version of the MSLQ, presented to participants as the Aviation Training Effectiveness Questionnaire, was adapted to assess these factors within the context of the specific training environment and was administered as a post-test after the training to assess the differential impact of the query method, at the *state* level (i.e., task-specific), on participants' self-reports for these constructs. Reliability analysis revealed acceptable reliability coefficients for the self-efficacy and metacognition scales in both the pre-test ($\alpha = .931$ and $.726$, respectively) and post-test ($\alpha = .945$ and $.813$, respectively) modified versions of the MSLQ.

For both versions of this paper-and-pencil questionnaire, participants were presented with 20 statements (8 self-efficacy items and 12 metacognition items), and asked to circle the number that best described the way they felt concerning that statement. Participants' responses were recorded on a 7-point Likert-type scale, ranging from 1 (*not at all true of me*) to 4 (*somewhat true of me*) to 7 (*very true of me*). An illustrative example of a self-efficacy statement in the pre-test (trait-level/general predisposition) modified version of the MSLQ would be: "I'm confident I can learn the basic concepts taught in my classes." For the post-test (state-level/task-specific), this self-efficacy statement would read: "I'm confident I learned the basic concepts taught in the Aviation Training Tutorial." An illustrative example of a metacognitive self-regulation statement in the pre-test (trait-level/general predisposition) would be: "When I become confused about something I'm reading for my classes, I go back and try to figure it out." For the post-test (state-level/task-specific), this metacognitive self-regulation statement would read: "When I became confused about something I was reading in the Aviation Training Tutorial, I went back and tried to figure it out."

Verbal Comprehension Ability

The nature of the material presented in the Aviation Training Tutorial required the understanding and integration of complex concepts and relations. Furthermore, as noted earlier, verbal comprehension ability has been shown to be positively related to complex skill acquisition as well as metacomprehension ability. As was utilized in the Cuevas et al. (2002) study investigating the differential impact of diagrammatic presentation for learners varying in verbal comprehension ability, Part 1 (Verbal Comprehension) of the Guilford-Zimmerman Aptitude Survey (Guilford & Zimmerman, 1981) was administered to assess individual differences in this

learner aptitude. Part 1 has a computed odd-even estimate of reliability of .96 (Guilford & Zimmerman, 1981). Factorial validity is demonstrated from the results of three factor analyses, with factor loadings ranging from .70 to .86 (Guilford & Zimmerman, 1981). For this paper-and-pencil task, participants were given ten minutes to respond to 72 multiple-choice questions assessing knowledge of semantic meanings. Specifically, participants were presented with a word (e.g., *earth*) and were then asked to select from a list of five other words (e.g., *sugar, farm, sun, soil, horse*) the one of these that had a meaning like the first word (e.g., *soil*).

Time-on-Task

Using a chronometer (i.e., digital stopwatch), an experimenter observed and recorded time-on-task for all participants as they completed both the training and performance portions of the experiment. Specifically, time-on-task (measured in minutes) was recorded for completion of the Aviation Training Tutorial (i.e., training) and the card sort and knowledge assessment tasks (i.e., performance).

Apparatus

The software program for the Aviation Training Tutorial, Knowledge Assessment Task, card sort tutorial, and card sort program were hosted on an IBM compatible Pentium 586 computer with a 15-inch color monitor, run on Windows XP[®] operating system. The Aviation Training Tutorial, Knowledge Assessment Task, and card sort tutorial were presented utilizing Microsoft PowerPoint XP[®]. Participants navigated through the tutorials and test using a standard point-and-click mouse. Hyperlinks were incorporated into the presentation of the tutorial material. Multimedia presentation using audio-video-interleaved (AVI) files were incorporated

into the Airplane Function Assessment (integrative knowledge questions). The TPL-KATS-card sort program was written in the Java[®] programming language. Java 1.2[®] virtual machine was installed on the computer in order to operate the software (for more information, see <http://java.sun.com>). Both keyboard inputs and the use of a standard point-and-click mouse were required to perform the card sort task. A paper-and-pencil format was used to record participants' responses to the queries embedded in the training, metacomprehension prediction and postdiction assessments, Knowledge Assessment Task, verbal comprehension ability measure, and the pre- and post-test modified versions of the MSLQ.

Procedure

Upon arrival, participants were randomly assigned to one of the three experimental groups. The experiment consisted of two parts. In the first half of the experiment, participants completed an informed consent form (see Appendix A), biographical data form (e.g., age, gender, year in school, prior knowledge of aviation – see Appendix D), the modified pre-test version of MSLQ assessing general (i.e., trait-level) responses to the self-efficacy and metacognitive self-regulation scales, and the measure of verbal comprehension ability. In the second half of the experiment, participants received computer-based instruction on the principles of flight using the Aviation Training Tutorial, and proceeded with self-paced instruction through the tutorial. For the query conditions, the sentence task was embedded immediately following the last lesson presented in each module of the tutorial.

Upon completion of each module, all participants were asked to complete the Module Survey. All participants then proceeded to the next module in the tutorial. The participants continued this sequence until all three modules were completed. At the completion of the

tutorial, participants were then asked to complete the Tutorial Survey. Participants were then presented with the card sort tutorial and card sort task. After the card sort task, all participants were asked to complete the Aviation Training Effectiveness Questionnaire, that is, the task-specific modified version of the MSLQ self-efficacy and metacognitive self-regulation scales, which served as the post-test. Participants were then presented with the Knowledge Assessment Task, followed by the Knowledge Assessment Questionnaire. Finally, participants completed a study survey, were debriefed, and extra credit was assigned. On average, the total length of the experiment, including training and performance assessment was approximately ninety-five minutes.

RESULTS

Analyses

The experimental results were analyzed separately in terms of post-training cognitive (i.e., knowledge organization, knowledge acquisition, instructional efficiency) and metacognitive (i.e., metacomprehension accuracy, metacognitive self-regulation) outcomes. When deemed appropriate, correlations were calculated. An alpha level of .05 was used for all statistical analyses. Results of these analyses will be presented as follows. The analysis for the check of random assignment will be presented first. Contrary to predictions, individual differences in verbal comprehension ability did not interact with the query method to influence post-training outcomes. However, analysis did show that verbal comprehension ability was significantly correlated with several of the cognitive measures (refer to Table 1). Accordingly, the next two sections report the results of the analyses focusing on the main effect of the query method on post-training cognitive and metacognitive outcomes, respectively, with verbal comprehension ability treated as a covariate, as appropriate. The fourth section reports the results of the analysis evaluating the hypothesized differential effect of the query method on task-specific metacognitive self-regulation and self-efficacy expectations of performance. Finally, results of the content analysis of the sentences generated by the two query training conditions (LLEQ, HLEQ) will be presented.

Table 1 lists the intercorrelations for the cognitive and metacognitive measures and verbal comprehension ability (individual differences variable). Tables 2 and 3 report the unadjusted means and standard deviations of all relevant cognitive and metacognitive measures, respectively, for the three training conditions (NQ, LLEQ, HLEQ).

Table 1

Intercorrelations of Verbal Comprehension Ability and Cognitive and Metacognitive Measures^a

Dependent Variable	1	2	3	4	5	6	7	8	9
Verbal Comprehension	—	.05	.38**	.32*	.41**	.23	-.11	-.12	-.18
Knowledge Organization		—	.30*	.27	.39**	.06	-.19	-.21	-.10
Knowledge Assessment									
Total			—	.89**	.90**	.77**	-.65**	-.53**	.10
Declarative				—	.68**	.52**	-.68**	-.54**	.08
Perceptual					—	.61**	-.53**	-.47**	.03
Integrative						—	-.40**	-.31*	.18
Prediction Bias							—	.63**	-.20
Postdiction Bias								—	-.20
Level of Review Effort									—

^a $N = 51$.* $p < .05$ (two-tailed). ** $p < .01$ (two-tailed).

Table 2

Unadjusted Means and Standard Deviations for Cognitive Measures by Training Condition

Dependent Variable	Training Condition ^a		
	NQ	LLEQ	HLEQ
Knowledge Organization	.499 (.197)	.621 (.199)	.437 (.232)
Knowledge Assessment			
Total	.507 (.122)	.654 (.190)	.553 (.175)
Declarative	.532 (.172)	.650 (.230)	.565 (.170)
Perceptual	.556 (.151)	.696 (.192)	.578 (.189)
Integrative	.365 (.173)	.588 (.220)	.482 (.260)
Instructional Efficiency			
Declarative	-0.457 (0.668)	0.520 (0.937)	-0.065 (1.230)
Perceptual	-0.476 (0.676)	0.599 (0.944)	-0.125 (1.326)
Integrative	-0.618 (0.765)	0.605 (1.016)	0.011 (1.415)

Note. *Knowledge Organization* values represent mean correlation with expert model. Values for *Knowledge Assessment* represent mean percent correct. *Instructional Efficiency* values represent mean E score.

^a $n = 17$ for each condition.

Table 3

Unadjusted Means and Standard Deviations for Metacognitive Measures by Training Condition

Dependent Variable	Training Condition ^a		
	NQ	LLEQ	HLEQ
Metacomprehension Accuracy			
Level of Understanding	-.169	.302	.513
Prediction Bias	.246 (.169)	.104 (.225)	.159 (.162)
Postdiction Bias	.046 (.190)	.022 (.155)	.041 (.209)
Metacomprehension Calibration	.200 (.112)	.082 (.180)	.118 (.170)
Metacognitive Self-Regulation			
Level of Review Effort	1.941 (1.853)	5.294 (2.366)	5.235 (2.077)

Note. Values for *Level of Understanding* represent mean correlation with performance. Values for *Prediction* and *Postdiction Bias* represent percent difference from performance. Values for *Metacomprehension Calibration* represent percent difference between *Prediction* and *Postdiction Bias* scores. Values for *Level of Review Effort* represent mean score for overall review during training.

^a $n = 17$ for each condition.

Check of Random Assignment

To check the effectiveness of the procedure used to randomly assign participants to the experimental groups, univariate one-way ANOVAs were performed on variables that should not have been influenced by the different manipulations. These variables were: (a) verbal comprehension ability (as measured by the Guilford-Zimmerman Aptitude Survey Part 1), (b)

trait-level self-efficacy (as measured by the MSLQ pre-test), and (c) trait-level metacognitive self-regulation (as measured by the MSLQ pre-test). As indicated in Table 4, analysis revealed no significant differences among the three experimental groups on these variables.

Table 4

Means and Standard Deviations for Background Variables by Training Condition

Variable	Training Condition ^a			ANOVA Results	
	NQ	LLEQ	HLEQ	<i>F</i> (2, 48)	<i>p</i>
Verbal Comprehension	.373 (.116)	.367 (.137)	.338 (.118)	< 1	.691
Self-Efficacy	5.66 (1.043)	5.70 (0.677)	5.62 (0.983)	< 1	.967
Metacognitive					
Self-Regulation	5.11 (0.859)	4.66 (0.657)	4.90 (0.705)	1.60	.213

Note. Values for verbal comprehension ability represent mean percent correct. Responses to self-efficacy and metacognitive self-regulation scales measured on a 7-point scale, with responses ranging from 1 to 7.

^a *n* = 17 for each condition.

Effect of Query Method on Cognitive Outcomes

The cognitive measures were analyzed using a one-way between-groups MANCOVA, with query method serving as the independent variable and verbal comprehension ability treated as a covariate. The dependent measures included: knowledge organization, as measured via the card sort task, and knowledge acquisition, as measured using the three sets of knowledge assessment questions (i.e., declarative, perceptual, integrative). Multivariate analysis (reported

using Roy's largest root) revealed a significant effect for the query method on these cognitive products, $F(4,45) = 3.202, p = .021$. Verbal comprehension ability was a significant covariate, $F(4,44) = 2.733, p = .041$. Univariate analysis for each of the individual measures will be presented next. Adjusted means and standard errors are reported in Table 5.

Table 5

Adjusted Means and Standard Errors for Knowledge Organization and Knowledge Assessment Measures by Training Condition

Dependent Variable	Training Condition ^a			ANOVA Results	
	NQ	LLEQ	HLEQ	$F(2, 47)$	p
Knowledge Organization	.499 (.051)	.621 (.051) _a	.438 (.052) _b	3.269	.047
Knowledge Assessment					
Total	.500 (.037) _c	.650 (.037) _d	.564 (.037)	4.177	.021
Declarative	.526 (.045)	.646 (.045)	.575 (.045)	1.840	.170
Perceptual	.551 (.040) _a	.691 (.040) _b	.591 (.040)	3.343	.044
Integrative	.358 (.052) _c	.585 (.052) _d	.492 (.052)	4.742	.013

Note. Knowledge Organization values represent mean correlation with expert model. Values for Knowledge Assessment represent mean percent correct. Means in the same row with different subscripts 'a' and 'b' differ significantly at $p < .05$, two-tailed, by the Fisher Least Significant Difference (LSD) comparison; different subscripts 'c' and 'd' indicate significant difference at $p < .01$, two-tailed.

^a $n = 17$ for each condition.

Knowledge Organization – Card Sort Task

The card sort task was used to assess the degree to which presentation of the query method affected similarity to an expert model. A quantitative measure was derived from the card sort data to determine the connectedness among concepts. First, a list of all possible pairings of the 26 concepts was generated ($N = 325$). A value of 1 was assigned to pairings of concepts falling within the same group (i.e., if the participant grouped the pair of concepts together in the same category) and a value of 0 was assigned for the remaining concept pairs (i.e., for pairings where the participants did not group the two concepts together in the same category).

For this analysis, each participant's card sort data (i.e., the generated list of the participant's pairings of all the concepts) was compared to the card sort data generated by the subject matter expert. This expert had approximately 7,000 hours as a pilot and approximately 2,700 hours as an instructor and participated in the creation and evaluation of the tutorial. Specifically, by calculating the Pearson r correlation coefficient between the participant's card sort data and the card sort data generated by the subject matter expert, a participant's sensitivity to identifying the critical relations among the concepts can be evaluated. Hence, the similarity of their pairings to the expert's model would indicate the accuracy of the participant's connections among critical concepts (i.e., the participant's knowledge organization of the task). Although some researchers have raised concerns regarding the use of a single expert model to assess knowledge structure development (e.g., Shanteau, 1989), previous research has shown that experts do agree on structural relations generated from card sort tasks (Fiore, Fowlkes, Martin-Milham, & Oser, 2000) and that card sort data using a single expert model may be effectively used to ascertain the degree to which trainees accurately view conceptual relations (e.g., Fiore et al., 2003; Fiore et al., 2002; Scielzo et al., 2004).

Univariate tests revealed a significant effect of the query method on participants' knowledge organization, as indicated by the mean correlation between the participants' and the expert model's card sort pairings, $F(2,47) = 3.269, p = .047$. However, pairwise comparisons (using Fisher's Least Significant Difference) showed an unexpected pattern of results. Specifically, the LLEQ participants exhibited a significantly greater mean correlation with the expert model than the HLEQ participants, $p = .016$. Although the LLEQ participants' card sort mean correlation to the expert model was also greater than the NQ participants, this difference was not significant, $p = .098$ (refer to Table 5). No significant difference was found between the NQ and the HLEQ participants, $p = .413$. Thus, only limited support was found for the hypothesis involving the effect of the query method on participants' knowledge organization, favoring the LLEQ training condition (*Hyp1*).

Knowledge Assessment – Overall Performance

Univariate tests revealed a significant effect of the query method on participants' performance on the knowledge assessment task overall, $F(2,47) = 4.177, p = .021$. Specifically, pairwise comparisons (using Fisher's Least Significant Difference) showed that the LLEQ participants significantly outperformed the NQ participants on the knowledge assessment task overall, $p = .006$. Although the LLEQ participants' performance was also greater than the HLEQ participants, this difference was not significant, $p = .106$ (refer to Table 5). No significant difference was found between the NQ and the HLEQ participants, $p = .230$.

Knowledge Assessment – Declarative and Perceptual Knowledge

As hypothesized, univariate tests revealed no significant difference in performance on the declarative knowledge assessment questions, $F(2,47) = 1.840, p = .170$. Although the LLEQ participants did perform better than the NQ participants and HLEQ participants, these differences were not significant, $p > .05$ (refer to Table 5). These results support the hypothesis regarding the predicted lesser effect of the query method on acquisition of declarative knowledge (*Hyp2B*).

Contrary to predictions, univariate tests revealed a significant difference in performance on the perceptual knowledge assessment questions, $F(2,47) = 3.343, p = .044$. Specifically, pairwise comparisons (using Fisher's Least Significant Difference) showed that the LLEQ participants significantly outperformed the NQ participants on the perceptual knowledge assessment questions, $p = .040$. Although the LLEQ participants' performance was also greater than the HLEQ participants, this difference was not significant, $p = .081$ (refer to Table 5). No significant difference was found between the NQ and the HLEQ participants, $p = .476$. These results, thus, do not support the hypothesis regarding the predicted lesser effect of the query method on acquisition of perceptual knowledge (*Hyp2B*).

Knowledge Assessment – Integrative Knowledge

As hypothesized, univariate tests revealed a significant difference in performance on the integrative knowledge assessment questions, $F(2,47) = 4.742, p = .013$. Specifically, pairwise comparisons (using Fisher's Least Significant Difference) showed that the LLEQ participants significantly outperformed the NQ participants on the integrative knowledge assessment questions, $p = .004$. Although the LLEQ participants' performance was also greater than the

HLEQ participants, this difference was not significant, $p = .219$ (refer to Table 5). No significant difference was found between the NQ and the HLEQ participants, $p = .078$. Thus, these results support the hypothesis regarding the effect of the query method on acquisition of integrative knowledge, but only for the LLEQ training condition (*Hyp2A*).

Instructional Efficiency

The subjective workload (cognitive load) associated with learning the instructional material was determined by asking participants to report how easy or difficult they found it to understand the concepts presented in the tutorial, with responses recorded on a 7-point scale ranging from 1 (*very easy*) to 7 (*very difficult*). Table 6 reports the means and standard errors for subjective workload and the instructional efficiency scores associated with the different training conditions. A one-way between-groups ANOVA was conducted, comparing self-reported workload for participants in the three training conditions. Univariate tests revealed no significant differences in perceived workload during training (based on responses given on the Tutorial Survey following completion of the training), $F(2,48) = 2.768$, $p = .073$ (refer to Table 6).

Next, as in the Cuevas et al. (2002) study investigating the differential impact of diagrammatic presentation on post-training outcomes, the instructional efficiency (E) of the training program was calculated by plotting the standardized scores on measures of mental effort (R) (i.e., subjective report of task difficulty as indicated by responses on the Tutorial Survey) against the standardized scores on measures of performance (P) (i.e., declarative, perceptual, integrative), displayed as a cross of axes (for a detailed description of this procedure, see Paas & Van Merriënboer, 1993). Instructional efficiency was calculated using the following equation (adapted from Kalyuga, Chandler, & Sweller, 1999): $E = (P - R) / \text{SQRT}(2)$. The values of P

and R determine the sign of E. If $P > R$, then E will be positive, indicating higher efficiency (i.e., mental effort exerted is less, relative to the standard effort required to achieve that level of performance). If $P < R$, then E will be negative, indicating lower efficiency (i.e., mental effort exerted is greater, relative to the standard effort required to achieve that level of performance). Baseline (or standard level of efficiency) is represented by $E = 0$.

Table 6

Means and Standard Errors for Subjective Workload and Instructional Efficiency Scores by Training Condition

Dependent Variable	Training Conditions ^a			ANOVA Results	
	NQ	LLEQ	HLEQ	<i>F</i> (2, 48)	<i>p</i>
Subjective Workload	3.412 (0.283)	2.471 (0.283)	2.941 (0.283)	2.768	.073
Instructional Efficiency					
Declarative	-0.457 (0.236) _c	0.520 (0.236) _d	-0.065 (0.236)	4.346	.018
Perceptual	-0.476 (0.247) _c	0.599 (0.247) _{ad}	-0.125 (0.247) _b	4.937	.011
Integrative	-0.618 (0.266) _c	0.605 (0.266) _d	0.011 (0.266)	5.272	.009

Note. *Instructional Efficiency* values represent mean E score. Means in the same row with different subscripts ‘a’ and ‘b’ differ significantly at $p < .05$, two-tailed, by the Fisher Least Significant Difference (LSD) comparison; different subscripts ‘c’ and ‘d’ indicate significant difference at $p < .01$, two-tailed.

^a $n = 17$ for each condition

Instructional efficiency (E) scores were analyzed using a one-way between-groups MANOVA, with query method serving as the independent variable (note: verbal comprehension ability was not a significant covariate, and thus, was not included in the analyses). The dependent measures reflected the instructional efficiency of the training program in relation to performance on the three sets of knowledge assessment questions (i.e., declarative, perceptual, integrative).

As hypothesized, multivariate analysis (reported using Roy's largest root) revealed a significant effect for the query method on the training program's instructional efficiency, $F(3,47) = 3.648, p = .019$. Univariate tests revealed a significant effect of the query method on the training program's instructional efficiency in relation to performance on all three sets of knowledge assessment questions: declarative, $F(2,48) = 4.346, p = .018$; perceptual, $F(2,48) = 4.937, p = .011$; and, integrative, $F(2,48) = 5.272, p = .009$.

Overall, the LLEQ training condition consistently yielded positive instructional efficiency scores (i.e., greater performance was achieved with less perceived cognitive effort), whereas the NQ training condition consistently yielded negative instructional efficiency scores (i.e., poorer performance was achieved with greater perceived cognitive effort) (refer to Figure 9). Instructional efficiency for the HLEQ training condition was typically at baseline (near zero) (i.e., standard level of performance was achieved relative to perceived cognitive effort). Pairwise comparisons (using Fisher's Least Significant Difference) showed that, in general, instructional efficiency scores across the three knowledge assessment questions were significantly different between the NQ and LLEQ training conditions, but not significantly different between the LLEQ and HLEQ training conditions or between the NQ and HLEQ training conditions (refer to Table 6). Thus, these results support the hypothesis regarding the

beneficial effect of the query method on the training program’s instructional efficiency, but only for the LLEQ training condition (*Hyp3*).

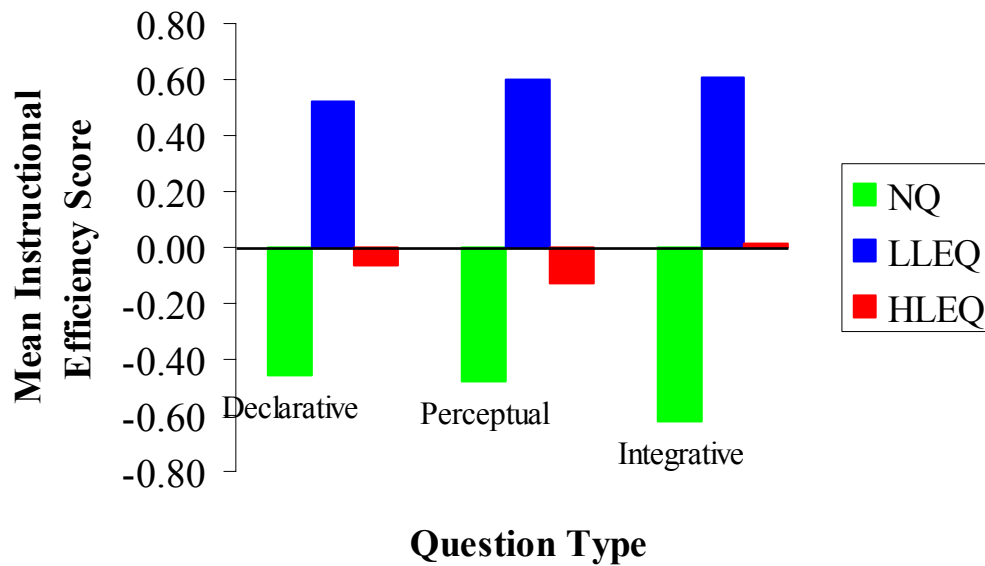


Figure 9. Effect of query method on training program’s instructional efficiency.

Effect of Query Method on Metacognitive Outcomes

Separate analyses were conducted to examine the effect of the query method on metacognitive outcomes, including: metacomprehension accuracy based on level of understanding; metacomprehension accuracy as measured using metacomprehension prediction and postdiction bias scores; metacomprehension calibration as indicated by mean differences between metacomprehension prediction and postdiction bias scores; and, metacognitive self-regulation as indicated by overall level of review effort. Verbal comprehension ability was not a significant covariate on the metacognitive measures, and thus, was not included in the analyses.

Finally, the hypothesized differential effect of the query method on task-specific (i.e., state-level) metacognitive self-regulation and self-efficacy expectations of performance was also evaluated. Each of these analyses will be described in turn.

Metacomprehension Accuracy – Level of Understanding

The first measure of metacomprehension accuracy involved calculating the Pearson r correlation coefficient between participants' subjective level of understanding of the concepts presented in the tutorial (based on their responses to the Tutorial Survey) and their overall performance on the knowledge assessment task. In general, participants' perceived level of understanding was significantly correlated with overall performance on the knowledge assessment task, $r(51) = .279, p = .048$, two-tailed. However, when analyzed by training condition, this correlation was significant only for the HLEQ participants, $r(17) = .513, p = .035$, two-tailed. Correlations for the NQ ($r(17) = -.169, p = .517$, two-tailed) and LLEQ ($r(17) = .302, p = .238$, two-tailed) participants were not significant. Thus, these results support the hypothesis regarding the effect of the query method on participants' metacomprehension accuracy based upon self-reported level of understanding, but only for the HLEQ training condition (*Hyp5B*).

Metacomprehension Accuracy – Bias Scores

For the second measure of metacomprehension accuracy, the effect of the query method on participants' overall metacomprehension prediction and postdiction bias scores (based upon responses to the Tutorial Survey) was analyzed using a one-way between-groups MANOVA, with query method serving as the independent variable. Multivariate analysis (reported using

Roy's largest root) revealed a significant effect for the query method on these metacognitive measures, $F(2,48) = 3.473$, $p = .039$. Means and standard errors are reported in Table 7. Univariate analysis for each bias score measure will be presented next.

Table 7

Means and Standard Errors for Metacomprehension Bias Scores by Training Condition

Dependent Variable	Training Conditions ^a			ANOVA Results	
	NQ	LLEQ	HLEQ	$F(2, 48)$	p
Prediction Bias	.246 (.045) _a	.104 (.045) _b	.159 (.045)	2.456	.096
Postdiction Bias	.046 (.045)	.022 (.045)	.041 (.045)	< 1	.926

Note. Values for *Prediction* and *Postdiction Bias* represent percent difference from performance. Means in the same row with different subscripts 'a' and 'b' differ significantly at $p < .05$, two-tailed, by the Fisher Least Significant Difference (LSD) comparison.

^a $n = 17$ for each condition.

Univariate analysis for the effects of the query method on metacomprehension *prediction* bias failed to achieve the established significance criterion, $F(2,48) = 2.456$, $p = .096$. Nevertheless, pairwise comparisons (using Fisher's Least Significant Difference) showed that prediction bias scores for the LLEQ participants were significantly lower than the prediction bias scores for the NQ participants, $p = .033$, indicating greater metacomprehension accuracy. Although the LLEQ participants' prediction bias scores were also smaller than the HLEQ participants, this difference was not significant, $p = .339$ (refer to Table 7). No significant

difference was found between the NQ and HLEQ participants, $p = .184$. Thus, these results provide limited support for the hypothesis regarding the effect of the query method on metacomprehension accuracy as indicated by lower prediction bias scores, but only for the LLEQ training condition (*Hyp5A*).

Univariate analysis for the effects of the query method on metacomprehension *postdiction* bias was not significant, $F < 1$ (refer to Table 7). It appears that all participants were better at estimating their performance following completion of the knowledge assessment task (cf. Fiore et al., 2002; Schwartz & Metcalfe, 1994). These results will be examined in more detail next.

Metacomprehension Calibration

To test the hypothesis that completion of the knowledge assessment task may assist participants in calibrating their metacomprehension, paired samples *t*-tests were used to analyze the difference between metacomprehension prediction and postdiction bias scores for the training conditions. Overall, prediction bias scores ($M = .1697$, $SE = .027$) were significantly greater than postdiction bias scores ($M = .0364$, $SE = .026$), $t(50) = 5.886$, $p < .0005$, two-tailed, across the three training conditions.

However, when examined by condition, this difference was significant only for NQ participants, $t(16) = 7.376$, $p < .0005$, two-tailed, and the HLEQ participants, $t(16) = 2.846$, $p = .012$, two-tailed. The difference between prediction and postdiction bias scores was not significantly different for the LLEQ participants, $t(16) = 1.884$, $p = .078$, two-tailed. Nevertheless, these results support the hypothesis that, overall, participants would exhibit significantly lower bias (indicating greater metacomprehension accuracy) following completion

of the knowledge assessment task (metacomprehension *postdiction*) as compared to their bias scores based on self-evaluations prior to assessment (metacomprehension *prediction*) (*Hyp6*).

Metacognitive Self-Regulation – Level of Review Effort

Incorporating the query method into the training was hypothesized to positively affect not only participants' *knowledge* of their cognitions (metacomprehension accuracy), but also the *regulation* of their cognitions (level of review effort). To test this hypothesis, a one-way between-groups MANOVA was conducted to first determine if there were any differences among the training conditions in time-on-task for completion of the training and performance portions of the experiment. The query method served as the independent variable. Dependent variables included time-on-task (measured in minutes) for completing the Aviation Training Tutorial (i.e., training) and the card sort and knowledge assessment tasks (i.e., performance).

Multivariate analysis (reported using Roy's largest root) revealed a significant effect for the query method on time-on-task, $F(3,47) = 14.899, p < .0005$. However, univariate tests revealed a significant difference in time-on-task only for completion of the Aviation Training Tutorial, $F(2,48) = 15.111, p < .0005$ (refer to Table 8). Pairwise comparisons (using Fisher's Least Significant Difference) showed that both the LLEQ and the HLEQ participants, invested significantly greater time-on-task for their training than the NQ participants, $p < .0005$. No significant difference for training time-on-task was found between the LLEQ and HLEQ participants, $p = .190$. Univariate tests revealed no significant difference for time-on-task in performing either the card sort task ($F < 1$) or the knowledge assessment task ($F < 1$). Overall, these findings indicate that the query method may have prompted participants to invest more

time in their training. Yet, the LLEQ and HLEQ participants did not necessarily need to take any longer than the NQ participants in performing the post-training tasks.

Table 8

Means and Standard Errors for Time-On-Task and Level of Review Effort by Training Condition

Dependent Variable	Training Conditions ^a			ANOVA Results	
	NQ	LLEQ	HLEQ	<i>F</i> (2, 48)	<i>p</i>
Time on Task					
Tutorial	17.824 (1.378) _a	25.529 (1.378) _b	28.118 (1.378) _b	15.111	<.0005
Card Sort	18.765 (1.225)	20.529 (1.225)	18.588 (1.225)	< 1	.470
Knowledge Assessment	18.706 (1.231)	20.235 (1.231)	21.000 (1.231)	< 1	.413
Level of Review Effort	1.941 (0.512) _a	5.294 (0.512) _b	5.235 (0.512) _b	14.074	<.0005

Note. Values for *Time on Task* represent mean time in minutes. Values for *Level of Review Effort* represent mean score for overall review during training. Means in the same row with different subscripts ‘a’ and ‘b’ differ significantly at $p < .01$, two-tailed, by the Fisher Least Significant Difference (LSD) comparison.

^a $n = 17$ for each condition.

One could argue that the extra time invested in training by the LLEQ and HLEQ participants was simply a byproduct of their training condition (i.e., completing the queries added to training time-on-task). Therefore, it is important to evaluate how effectively this additional time-on-task during training was utilized. To address this question, a one-way between-groups ANOVA was conducted, with query method as the between-groups variable and

overall observed level of review effort as the dependent variable. Univariate tests revealed a significant effect for the query method on participants' observed level of review effort, $F(2,48) = 14.074$, $p < .0005$. Pairwise comparisons showed that both the LLEQ and the HLEQ participants exhibited significantly greater levels of review effort during their training than the NQ participants, $p < .0005$ (refer to Table 8). No significant difference for level of review effort was found between the LLEQ and HLEQ participants, $p = .936$. Thus, these findings support the hypothesis that presentation of the query method would lead to significantly greater observed levels of review effort among the participants (*Hyp7*).

Metacognition versus Self-Efficacy

Consistent with previous studies (cf. Ford et al., 1998), analysis revealed a significant correlation between metacognitive self-regulation and self-efficacy expectations of performance, as indicated by responses to the MSLQ pre-test, $r(51) = .481$, $p < .0005$, two-tailed, as well as by responses to the MSLQ post-test, $r(51) = .385$, $p = .005$, two-tailed. To assess the differential impact of the query method on participants' self-reports of task-specific (i.e., state-level) metacognitive self-regulation and self-efficacy expectations of performance, a one-way between-groups MANCOVA was conducted, with the query method serving as the between-groups variable. MSLQ pre-test scores on these constructs were incorporated into the analysis as covariates to account for pre-existing (i.e., trait-level) within-group variance on metacognitive self-regulation and self-efficacy expectations of performance.

Contrary to predictions, multivariate tests (reported using Roy's largest root) failed to achieve the established significance criterion, $F(2,46) = 2.914$, $p = .064$. Specifically, although the pattern of results (refer to Table 9) was consistent with the data reported for the other

metacognitive measures, after accounting for the variance attributable to these factors at the trait level, the query method did not have the hypothesized differential effect on participants' self-reports of task-specific (i.e., state-level) metacognitive self-regulation and self-efficacy expectations of post-training performance (*Hyp9A* and *Hyp9B*).

Table 9

Means and Standard Errors for MSLQ Post-Test Measures by Training Condition

Dependent Variable	Training Conditions ^a			ANOVA Results	
	NQ	LLEQ	HLEQ	<i>F</i> (2, 46)	<i>p</i>
Self-Efficacy	4.212 (0.197)	4.855 (0.198)	4.567 (0.193)	2.552	.089
Metacognitive					
Self-Regulation	4.004 (0.190)	4.397 (0.191)	4.531 (0.186)	2.087	.136

Note. Responses to self-efficacy and metacognitive self-regulation scales measured on a 7-point scale, with responses ranging from 1 to 7. Values represent mean response for each scale.

^a *n* = 17 for each condition.

Content Analysis of Participants' Queries

A content analysis of the participants' responses to the queries was conducted to determine if there were any significant qualitative differences between the LLEQ and HLEQ training conditions. Specifically, the sentences generated by the participants in the LLEQ and HLEQ training conditions were evaluated with regard to the number of concepts used in each sentence as well as the accuracy of the sentences. Recall the HLEQ training condition prompted

participants to generate sentences using *three* or more concepts from the list presented whereas the LLEQ training condition prompted participants to generate sentences using only *one* of these concepts. The number of concepts used in the sentence generated for each of the three modules was summed to calculate the total number of concepts used in the sentence generation task during training. The expected minimum number of concepts possible would be three for the LLEQ training condition (i.e., one concept per module) and nine for the HLEQ training condition (i.e., at least three concepts per module). Accuracy of the sentence generated in each module was rated as either “0” for inaccurate or “1” for accurate, resulting in a range from 0 (i.e., none of the sentences were accurate) to 3 (i.e., all three sentences were accurate).

Separate independent samples *t*-tests were conducted to analyze the total number of concepts and sentence accuracy, with training condition (LLEQ or HLEQ) serving as the between-groups factor. As would be expected, analysis showed that participants in the HLEQ training condition ($M = 10.940$; $SD = 2.384$) used a significantly greater total number of concepts in generating their sentences than participants in the LLEQ training condition ($M = 5.060$; $SD = 2.164$), $t(32) = 7.532$, $p < .0005$, two-tailed. However, with regard to the accuracy of the sentences, participants in the LLEQ training condition ($M = 2.820$; $SD = 0.529$) generated significantly more accurate sentences than participants in the HLEQ training condition ($M = 2.350$; $SD = 0.786$), $t(32) = 2.049$, $p = .049$, two-tailed.

Furthermore, analysis revealed that, overall, sentence accuracy was significantly positively correlated with performance on several of the cognitive measures: knowledge organization (i.e., mean correlation with the expert model), overall performance on the knowledge assessment task, and specific performance on the declarative and perceptual knowledge assessment questions (refer to Table 10). With regard to the metacognitive measures,

sentence accuracy was significantly negatively correlated to prediction bias, indicating that higher sentence accuracy was associated with lower bias scores (refer to Table 10).

Table 10

Correlations for Sentence Accuracy with Cognitive and Metacognitive Measures^a

Dependent Variable	<i>r</i>	<i>P</i> [*]
Knowledge Organization	.363	.017
Knowledge Assessment		
Total	.287	.050
Declarative	.287	.050
Recognition	.336	.026
Integrative	.088	.311
Level of Review Effort	.014	.469
Prediction Bias	-.310	.037
Postdiction Bias	-.178	.157

^a *N* = 34.

* one-tailed.

DISCUSSION

The present study explored the effectiveness of embedding a guided, learner-generated instructional strategy (query method), designed to support learners' cognitive and metacognitive processes, within the context of a computer-based complex task training environment. Additionally, this study also examined the effect of varying the level of elaboration prompted by the queries, asking participants to generate either simple (low level elaboration) or complex (high level elaboration) sentences. Overall, results consistently highlighted the beneficial effect of presenting participants with low-level elaboration queries, as compared to the no-query or high-level elaboration queries. These findings will be discussed next in greater detail.

Query Method – Low Level Elaboration

Incorporating the low-level elaboration queries into the training program significantly enhanced both cognitive (i.e., knowledge organization, knowledge acquisition, instructional efficiency) and metacognitive (i.e., metacomprehension accuracy, metacognitive self-regulation) post-training outcomes. In terms of cognitive outcomes, participants presented with the low-level elaboration queries exhibited significantly more accurate knowledge organization (as indicated by greater similarity to an expert model), better acquisition of perceptual knowledge, and superior performance on integrative knowledge assessment involving the integration and application of task-relevant concepts. Consistent with previous studies (cf., Cuevas et al., 2002; Fiore et al., 2003; Mayer & Gallini, 1990; Mayer & Sims, 1994), no significant differences in performance were found on basic factual knowledge assessment. Presentation of the low-level elaboration queries also significantly improved the training program's instructional efficiency, that is, greater performance was achieved with less perceived cognitive effort. Finally,

participants presented with the low-level elaboration queries generated significantly more accurate sentences than participants presented with the high-level elaboration queries.

In terms of metacognitive outcomes, participants presented with the low-level elaboration queries exhibited significantly greater metacomprehension accuracy (as indicated by significantly lower prediction bias scores) and more effective metacognitive self-regulation during training (as indicated by significantly greater observed levels of review effort). It should be also noted that the increased time on task during training exhibited by the participants presented with the low-level elaboration queries did not translate into perceptions of increased mental effort exerted during training. In other words, even though these participants invested significantly more time on task during their training, they did not perceive the training as being more difficult, as compared to the participants in the no-query training condition.

Moreover, this last finding involving observed level of review effort may explain the beneficial impact of the low-level elaboration queries on the perceptual knowledge assessment questions. Specifically, by investing significantly greater time-on-task during training and significantly increasing their level of review effort, the participants presented with the low-level elaboration queries may have increased their exposure to graphical illustrations of the key concepts in the training. This additional exposure may be one plausible hypothesis for explaining the significant differences in performance on the perceptual knowledge assessment questions between participants in the no-query and low-level elaboration query conditions. Perceptual knowledge can be defined as “the veridical and largely unverbalizable representation that develops via multiple exposures to environmental stimuli” (Fiore, Jentsch, Oser, & Cannon-Bowers, 2000, p. 17). Thus, by prompting participants to actively review the graphical illustrations of the key concepts presented in the training and increasing exposure to this critical

perceptual information, the query method may have facilitated perceptual learning (i.e. the acquisition of perceptual knowledge), leading to improved performance on the perceptual knowledge assessment questions.

Query Method – High Level Elaboration

Incorporating the high-level elaboration queries into the training consistently failed, with only a few exceptions, to produce significantly better post-training outcomes than the no-query or the low-level elaboration query training conditions. For example, by prompting participants to build internal associations between three or more concepts in the to-be-learned material, the high-level elaboration queries were expected to facilitate the development of a deeper, more integrative understanding of the information (cf., Weinstein & Mayer, 1986), leading to more accurate knowledge organization. However, participants presented with the low-level elaboration queries exhibited significantly more accurate knowledge organization than those presented with the high-level elaboration queries. Additionally, no significant differences in knowledge organization accuracy were found between the no-query and high-level elaboration query training conditions.

In fact, no significant differences were found between the no-query and high-level elaboration query training conditions on either knowledge acquisition (i.e., declarative, perceptual, and integrative knowledge assessment) or instructional efficiency. In other words, with regard to the cognitive measures, generating complex sentences was no more effective than not generating any sentences at all. Although the high-level elaboration queries may have, to some extent, enhanced participants' metacomprehension accuracy (as indicated by the significant correlation between perceived level of understanding and performance on the knowledge

assessment task) as well as their metacognitive self-regulation (as indicated by the significantly greater observed levels of review effort when compared to the no-query condition), this beneficial effect failed to translate to successful post-training cognitive outcomes (i.e., knowledge organization, knowledge acquisition).

For example, although the increased exposure to critical perceptual information was also evident for participants in the high-level elaboration query training condition, as would be expected given their significantly greater observed levels of review effort, these participants did not perform significantly better than the no-query participants on the perceptual knowledge questions. The beneficial effects of the query method on perceptual knowledge acquisition may have been diminished by the increased cognitive load associated with the complexity of completing the high-level elaboration sentence generation task. Specifically, the increased cognitive processing of the training material associated with the high-level elaboration queries may have imposed too great a cognitive load on participants during their training, minimizing the cognitive resources available for achieving greater learning and higher levels of performance on the cognitive measures, as was evident with the low-level elaboration training condition.

Metacognition versus Self-Efficacy

Given its learning-oriented versus performance-oriented design, the query method was expected to have a stronger positive effect on participants' self-reports of task-specific (i.e., state-level) metacognitive self-regulation during training than on their self-reports of task-specific (i.e., state-level) self-efficacy expectations of post-training performance. However, after accounting for the variance attributable to these factors at the trait level, the query method did

not have a significant differential effect on participants' task-specific metacognitive self-regulation and self-efficacy expectations of post-training performance.

Individual Differences – Verbal Comprehension Ability

A growing body of research has shown how training interventions may interact with differences in learner characteristics to produce differential results in learning (for a review, see Jonassen & Grawboski, 1993; Snow, 1997). As such, another objective of this dissertation was to examine how the effect of the query method would be differentially influenced by individual differences in verbal comprehension ability. Contrary to predictions, results revealed no significant interaction effects between verbal comprehension ability and query method training condition on either the cognitive or metacognitive measures.

Limitations and Implications for Future Research

Several limitations to the present study should be noted. First, the participants used in this study were undergraduate students enrolled in psychology courses. Consequently, external validity of this study is low and this limits the generalizability of these findings to complex operational environments. Future research needs to explore the utility of the query method within computer-based training for more operationally-valid populations. Further, the training material used in this study was based on introductory concepts related to the principles of flight. Although this domain is more complex relative to the training material explored in prior studies (e.g., mechanical instruments; cf. Mayer & Gallini, 1990; Mayer & Sims, 1994), to increase the external validity of this study's findings, the query method needs to be investigated with higher-level concepts of increasing complexity that are more relevant to advanced training programs.

Second, although the reliability of the overall knowledge assessment task was acceptable, the internal consistency of the individual sections was not as high as would be desired to draw definitive conclusions from these results. This decrease may be due to the limited number of questions in each section; thus, further testing with more items may resolve this issue. Still, it may be necessary to increase the degree of intercorrelation among the items such that each section is validly and reliably measuring the same construct (e.g., perceptual knowledge).

Third, although the sample size used in this study was large enough to evaluate the main effect of the query method on training outcomes, this may not have been sufficient to detect the predicted interaction between verbal comprehension ability and training condition. Given the strong positive relationship found between verbal comprehension ability and several of the cognitive measures, further research using larger samples is warranted to explore this potential aptitude-treatment interaction.

Finally, the somewhat artificial nature of this study coupled with its limited duration (i.e., brief training period, followed immediately by assessment) may not have been sufficiently robust to detect the differential effect of the query method on learners' self-reports of metacognitive self-regulation during training and self-efficacy expectations of post-training performance. In light of the significant positive correlations found between self-efficacy expectations of performance and metacognitive self-regulation (as indicated by responses to the MSLQ pre- and post-test measures), further research is clearly warranted with more realistic training programs to tease apart the complex relationship between these two constructs in order to better understand how these factors influence post-training outcomes.

TRANSFORMING THE LEARNING PROCESS: THEORETICAL AND PRACTICAL IMPLICATIONS

The overall goal of this dissertation was to further increase our understanding of the cognitive and metacognitive processes involved in learning within complex training environments and to investigate how instructional strategies can support this learning process. Furthermore, the results reported here build upon the findings of a program of research by Fiore and colleagues (Cuevas et al., 2002; Fiore et al., 2003; Fiore et al., 2002; Scielzo et al., 2004; for a review, see Cuevas et al., 2004) aimed at investigating the use of interactive computer-based training technology to facilitate knowledge acquisition and integration for complex task training environments.

Past research efforts in this area by Fiore and colleagues have explored the factors underlying knowledge acquisition and integration within two unique contexts: 1) the effect of diagrammatic presentations on the acquisition of domain knowledge in aviation (Cuevas et al., 2002; Fiore et al., 2003), and, 2) knowledge integration and cognitively diagnostic assessment within distributed team training environments (Fiore et al., 2002; Scielzo et al., 2004). Successive iterations of this investigation have examined not only knowledge acquisition and integration, but also the cognitive precursors to the development of this knowledge, including constructs such as metacomprehension (e.g., Maki et al., 1994) and instructional efficiency (e.g., Paas & Van Merriënboer, 1993) as well as the role of individual differences in learner aptitudes (e.g., Snow, 1997), to better converge on an understanding of technology-enhanced learning. This final section, therefore, attempts to integrate the significant findings in the present study within the consistent body of knowledge gathered during this multi-year programmatic research

effort in order to highlight both theoretical and practical implications for transforming the learning process into a constructive cognitive and metacognitive activity.

Theoretical Implications

Earlier work by Fiore and colleagues (Cuevas et al., 2002; Fiore et al., 2003) demonstrated how diagrammatic presentation can be effectively used to facilitate knowledge acquisition of complex systems (i.e., aviation) in a computer-based training environment. Specifically, embedding diagrams within complex task training may provide learners with a supporting framework for knowledge construction, enabling them to integrate the concepts presented in the training more effectively. The present study continues with this line of research by investigating how learner-generated elaboration of concepts may also support knowledge integration of a complex domain.

Learning as a Constructive Cognitive Activity

The results of the present study support the utility of *embedded-content dependent strategies* within complex task training environments. *Embedding* the low-level elaboration queries within the training may have prompted participants to attend to and interact with the critical concepts in the presented material, increasing the efficiency of their information processing and facilitating their comprehension monitoring and knowledge acquisition. The *content-dependent* nature of the queries manipulated in this study may have promoted learning of the target domain by emphasizing the key concepts and their unique interrelations, resulting in more accurate knowledge organization and better integration of task-relevant concepts. In general, these results, coupled with previous findings on the benefits of diagrammatic

presentation, highlight the importance of providing learners with instructional strategies that support the cognitive processes underlying knowledge acquisition.

Elaboration, Learner Control, and Cognitive Load

Prompting a high level of elaboration did not result in a significant difference in post-training outcomes, when compared to no elaboration. Several potential explanations may be offered to explain this somewhat counter-intuitive finding. For example, one could argue that arbitrarily forcing learners to generate complex sentences may limit the level of control they can exercise as they attempt to acquire an understanding of the training concepts. In contrast, giving learners the option of generating simple sentences may allow them to freely choose the appropriate level of complexity required in their elaboration of the material. Learner control is essential if learners are to be encouraged to take an active *mindful* approach to learning, that is, make a deliberate and systematic cognitive effort to engage the material during the learning process (Brown & Ford, 2002).

Another explanation for this lack of effect for the high level elaboration query focuses on the cognitive load imposed on learners during their training. The findings in the present study suggest that there may exist an optimal level of elaboration necessary to achieve the desired learning gains. Indeed, the most optimal performance outcomes were yielded by prompting participants to engage in a low level of elaboration of the training material. No elaboration of the training material, as associated with the no-query training condition, yielded significantly lower levels of post-training performance on both the cognitive and metacognitive measures. Yet, requiring participants to generate a high level of elaboration of the training material did not yield a corresponding gain in post-training outcomes, particularly for the cognitive measures.

Moreover, performance typically was not significantly greater than the no-query training condition. This lack of significant effect for the high-level elaboration query training condition may have been due to the increased cognitive load associated with generating complex sentences.

Such findings are consistent with Sweller's (1994) *cognitive load theory*, which proposes that training materials and activities required of learners during training should be structured to *minimize* any avoidable load on learners' cognitive resources (e.g., working memory capacity) and *maximize* knowledge structure development. Cognitive load may be influenced by both the content of the training material (i.e., complexity and level of integration of the domain concepts) as well as how the training material is presented (e.g., inadequate presentation modes that impose split-attention and redundant display of information) (Sweller, 1999; Sweller & Chandler, 1994). For instance, Marcus, Cooper, and Sweller (1996) argue that "If multiple elements must be considered simultaneously because of high element interactivity, cognitive load may be high and understanding difficult" (p. 50). Similarly, forcing learners to attend to multiple sources of information simultaneously may overburden their working memory capacity, reducing the cognitive resources available for successful learning to occur (e.g., Kalyuga et al., 1999). Accordingly, from a theoretical perspective, the findings in the present study suggest that when attempting to master an already inherently complex domain, instructional strategies that *force* the integration of these concepts (e.g., requiring generation of complex sentences) may negatively interfere with learners' knowledge construction by overloading their limited cognitive resources.

Learning as a Constructive Metacognitive Activity

According to Mayer (1999), "The educational goal of instruction is not solely the acquisition of well-defined pieces of knowledge, but also to help the learner develop metacognitive and self-regulation skills for learning" (p. 564). The results of the present study, when combined with the findings from earlier research on diagrammatic presentation (Cuevas et al., 2002), provide further evidence for how instructional strategies can be successfully employed to effectively enhance learners' metacognitive processes by increasing the accuracy of their comprehension monitoring (knowledge of cognitions). In addition, this study builds upon this earlier work by demonstrating how eliciting learner-generated elaboration of the training material may also induce learners to engage in the appropriate metacognitive self-regulation behaviors (regulation of cognitions), such as deliberately and mindfully examining the training material as needed, that is, increasing their level of review effort (Sorensen et al., 2001). This accurate comprehension monitoring and active regulation of their learning may have, in turn, enhanced learners' cognitive processes, translating to successful knowledge organization and knowledge integration. These findings add to the ever growing body of research documenting the significance of metacognitive processes for successful post-training outcomes.

Practical Implications

Past programmatic research efforts by Fiore and colleagues (Cuevas et al., 2002; Fiore et al., 2003; Fiore et al., 2002; Scielzo et al., 2004) have consistently demonstrated the value of adopting a multi-faceted approach for evaluating post-training outcomes within complex task training environments. As in this previous work, this dissertation employed an automated, computer-based training and performance assessment system that incorporated differing

instructional strategies (i.e., query method) with tests designed to tap specific components of knowledge (i.e., perceptual, declarative, integrative). These current findings provide further validation for utilizing several distinct yet related approaches designed to evaluate the impact of instructional strategies, such as the query method, on learners' cognitive and metacognitive processes.

Assessing Cognitive Outcomes

Critical post-training cognitive outcomes include the level of accuracy of learners' knowledge organization, that is, the degree to which learners' knowledge structures exhibit similarity to an expert model. In addition to mastery of both basic factual and perceptual knowledge, successful training programs must also prepare learners to effectively apply their newly acquired knowledge to more complex situations than were experienced during their training (Ford et al., 1998). As such, post-training assessment also needs to include tasks, utilizing dynamic task-relevant scenarios, that provide opportunities for evaluating how well learners can integrate and apply these different knowledge components (Fiore et al., 2002). The beneficial effect of the low-level elaboration queries on learners' cognitive processes was better diagnosed via such a multi-faceted approach to knowledge assessment, including both measures of knowledge organization and knowledge integration.

Evaluating a training program's instructional efficiency may serve as yet another potentially diagnostic measure of training effectiveness. Combining subjective ratings of cognitive effort with performance scores may reveal useful information about the effectiveness of training programs in terms of the cognitive costs associated with complex task training over and above what would be found by using measures of cognitive effort or performance alone

(Paas & Van Merriënboer, 1993). The results of the present study suggest that requiring participants to generate complex sentences using three or more concepts from the training material (high-level elaboration query) may have inadvertently increased the cognitive load associated with the training, limiting the cognitive resources available for successful learning to occur. Although, overall, this training condition yielded standard levels of performance relative to the perceived mental effort exerted during training, such baseline (near zero) instructional efficiency scores are not as ideal as the significantly higher instructional efficiency scores yielded for the low-level elaboration query training condition. As such, from a practical perspective, the present study highlights the importance of evaluating a training program's instructional efficiency, as this measure may be more diagnostic in determining why seemingly useful instructional design features may not lead to the most optimal post-training outcomes.

Assessing Metacognitive Outcomes

With regard to evaluating the effect of instructional strategies on learners' metacognitive processes, measures need to examine both learners' *knowledge* of their cognitions (e.g., metacomprehension accuracy) as well as the *regulation* of their cognitions (e.g., level of review effort). Such measures would provide useful diagnostic information to trainers regarding how well trainees are monitoring, evaluating, and regulating their knowledge acquisition process. For example, calculating bias scores would reveal not only the discrepancy between trainees' perceived level of performance and actual performance, but would also indicate the degree of underconfidence or overconfidence in their self-assessments. Similarly, observing trainees' level of review effort may reveal whether trainees are investing a sufficient amount of time in reviewing the training material as well as how effectively trainees are utilizing their time-on-task

during training. Such insightful information would provide guidance to trainers in selecting the appropriate training interventions to remedy any resulting deficiencies in performance.

CONCLUSION

With regard to the design of computer-based training programs, the results of this dissertation suggest that prompting learners to generate low-level (i.e., simple) elaborations of the training material may lead to improved post-training cognitive and metacognitive outcomes. Specifically, in terms of cognitive outcomes, low level elaboration of training concepts led to more accurate knowledge organization, better acquisition of perceptual knowledge, superior performance on tasks involving integration and application of concepts, and higher instructional efficiency. Improvements in metacognitive outcomes included greater metacomprehension accuracy and more effective metacognitive self-regulation during training. With regard to training evaluation, these findings implicate the importance of utilizing a multi-faceted approach to assessing the effectiveness of training programs, including measures of learners' cognitive as well as metacognitive processes. Equally important is gauging the training program's instructional efficiency, that is, the evaluation of a training program's design must also consider the cognitive costs associated with the training, relative to the performance achieved.

As advances in instructional design and distance learning technology drive organizations to rely more than ever before on technology-mediated distance learning approaches, instructional systems designers and training personnel need to garner a better understanding of the cognitive and metacognitive processes underlying successful post-training performance, as well as how individual differences in learner aptitudes may influence these processes (Annett, 1989; Winne & Stockley, 1998). Since the effective use of such training formats depends primarily on learners monitoring and regulating their own subjective learning experience, it is essential that program designers develop training programs that employ useful instructional strategies, guided by theory and research, to prompt learners to actively monitor their knowledge acquisition and

mindfully engage the material during the learning process (Brown & Ford, 2002). In sum, the true goal of education should be to transform the learning experience into a constructive cognitive and metacognitive activity (Glaser & Baxter, 2000). The line of research presented in this dissertation represents a promising first step toward achieving this challenging yet attainable objective.

APPENDIX A:
IRB COMMITTEE APPROVAL LETTER AND
APPROVED STUDENT INFORMED CONSENT FORM



University of
**Central
Florida**

Office of Research

June 6, 2004

Steven M. Fiore, Ph.D.
Team Performance Laboratory
12424 Research Parkway, Room 408
Orlando, FL 32826

Dear Dr Fiore:

With reference to your protocol entitled, "Supporting Knowledge Acquisition through Enhanced Displays (SKATED)," I am enclosing for your records the approved, executed document of the UCFIRB Form you had submitted to our office.

Please be advised that this approval is given for one year. Should there be any addendums or administrative changes to the already approved protocol, they must also be submitted to the Board. Changes should not be initiated until written IRB approval is received. Adverse events should be reported to the IRB as they occur. Further, should there be a need to extend this protocol, a renewal form must be submitted for approval at least one month prior to the anniversary date of the most recent approval and is the responsibility of the investigator (UCF).

Should you have any questions, please do not hesitate to call me at 823-2901.

Please accept our best wishes for the success of your endeavors.

Cordially,
Chris Grayson
Institutional Review Board (IRB)
Copies: Dr. Clint Bowers
IRB File



THE UNIVERSITY OF CENTRAL FLORIDA
INSTITUTIONAL REVIEW BOARD (IRB)

IRB Committee Approval Form

PRINCIPAL INVESTIGATOR(S): Stephen Fiore

PROJECT TITLE: Supporting Knowledge Acquisition through Enhanced Displays (SKATED).

Committee Members:

Dr. Theodore Angelopoulos: _____
Ms. Sandra Browdy: _____
Dr. Jacqui Byers: _____
Dr. Ratna Chakrabarti: _____
Dr. Karen Dennis: _____
Dr. Barbara Fritzsche: _____
Dr. Robert Kennedy: _____
Dr. Gene Lee: _____
Ms. Gail McKinney: _____
Dr. Debra Reinhart: _____
Dr. Valerie Sims: _____

Contingent Approval
Dated: _____
 Final Approval
Dated: _____

Expedited
Dated: 4 Jun 2004

Exempt
Dated: _____

Signed:  **Chair, IRB**
Dr. Sophia Dziegielewski

NOTES FROM IRB CHAIR (IF APPLICABLE): _____

UCF IRB
APPROVED
DATE

4 June 2004

Student Informed Consent Form

NAME: _____

Identification Number: _____

Introduction to Study:

This research, "Supporting Knowledge Acquisition through Enhanced Displays," is being conducted by principal investigators Stephen M. Fiore and Clint A. Bowers.

In this research, you will participate in a training program targeted at transfer of training. The experiment will focus on training and use of displays with differing levels of augmentation and performance evaluation using various tasks. Performance on these tasks will remain completely confidential (see below). The experiment should take approximately two to three hours. Upon completion of the study, course credit for participation in an experiment will be given in accordance with the procedures established within the Department of Psychology.

Risks and Benefits:

Participation in the current study does not involve any risks other than those commonly associated with the use of computer display terminals. Potential benefits include the development of guidelines for the appropriate use of training materials in a variety of differing task contexts. If you are injured during this study, as a result of the negligence of the Principal Investigator, the University of Central Florida. The Board of Regents of the State of Florida and the State of Florida shall be liable to the limited extent required by Florida law. You may seek appropriate compensation for injury by contacting the Personal Injury Insurance Coordinator at University of Central Florida Office of the General Counsel, Administration Building, Suite 350, Orlando, FL 32816-0015. The telephone number is (407) 823-2482. All inquiries to the Personal Injury Insurance Coordinator must be made in writing via either U.S. Mail, e-mail (gcounsel@mail.ucf.edu), or facsimile: (407) 823-6155.

Confidentiality of Personal Data:

All data you contribute to this study will be held in strict confidentiality by the researchers and will be kept under lock and key; that is, your individual data will not be revealed to anyone other than the researchers and their immediate assistants.

To insure confidentiality, the following steps will be taken: (a) only researchers will have access to the data in paper or electronic form. Data will be stored in locked facilities; (b) the actual forms will not contain names or other personal information. Instead, the forms will be matched to each participant by a number assigned by and only known to the experimenters; (c) only group means scores and standard deviations, but not individual scores, will be published or reported.

YOUR PARTICIPATION IN THIS RESEARCH IS COMPLETELY VOLUNTARY. YOU MAY WITHDRAW FROM PARTICIPATION AT ANY TIME WITHOUT PENALTY - THIS INCLUDES REMOVAL/DELETION OF ANY DATA YOU MAY HAVE CONTRIBUTED. SHOULD YOU DECIDE

NOT TO COMPLETE THE TRAINING STUDY, HOWEVER, YOU WILL BE ELIGIBLE ONLY TO THE COURSE CREDIT FOR THAT PART OF THE STUDY YOU HAVE COMPLETED.

Consent and Signatures:

By signing this form I agree to participate in the study "Supporting Knowledge Acquisition through Enhanced Displays," conducted by principal investigators Stephen M. Fiore and Clint A. Bowers.

I have been given the opportunity to ask the research assistants any questions I may have. I have read the procedure described above. I voluntarily agree to participate in the procedure, and I have received a copy of this form. For any other questions regarding this research, I can contact: Dr. Stephen M. Fiore, Research Scientist, Team Performance Laboratory, University of Central Florida, Orlando, FL 32826. Phone: (407) 384-2098; E-mail: stlore@pegasus.cc.ucf.edu

Signature: _____

PI Signature: _____

Date: _____

Date: _____

Version #7; Last revised April 30,2004

APPENDIX B:
AVIATION TRAINING TUTORIAL SURVEY

Participant Number _____

Aviation Training Tutorial Survey

The following questionnaire is designed to inform us about the effectiveness of the Aviation Training Tutorial that you just completed. Please circle the number that best describes the way you feel concerning that question.

1) Overall, how *helpful* was the Aviation Training Tutorial in teaching you about aviation concepts?

1	2	3	4	5	6	7
NOT AT ALL HELPFUL			SOMEWHAT HELPFUL			VERY HELPFUL

2) Overall, how *easy or difficult* did you find it to understand the concepts presented in the tutorial?

1	2	3	4	5	6	7
VERY DIFFICULT			FAIRLY EASY			VERY EASY

3) Overall, what is your *level of understanding* of the material presented in the tutorial that you just completed?

1	2	3	4	5	6	7
VERY POOR			FAIRLY GOOD			VERY GOOD

4) Based on your level of understanding, how well would you do on multiple-choice questions that ask you about the material presented in the tutorial that you just completed?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
None Correct					Half Correct					All Correct

APPENDIX C:
KNOWLEDGE ASSESSMENT QUESTIONNAIRE

Knowledge Assessment Questionnaire

The following questionnaire is designed to inform us about how well the Aviation Training Tutorial prepared you to respond to the knowledge assessment questions that you just completed. Please circle the number that best describes the way you feel concerning that question.

1) How well do you think you did **overall** on all the Knowledge Assessment questions that you just completed?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
None Correct			Half Correct				All Correct			

2) How well do you think you did on the **Factual Knowledge Assessment** questions that you just completed (that is, the first set of questions: the ones without pictures)?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
None Correct			Half Correct				All Correct			

3) How well do you think you did on the **Airplane Function Assessment** questions that you just completed (that is, the second set of questions: the ones with the animated pictures of airplanes)?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
None Correct			Half Correct				All Correct			

4) How well do you think you did on the **Concept Recognition Assessment** questions that you just completed (that is, the third set of questions: the ones with pictures of airplane parts, axes, movements, and instruments)?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
None Correct			Half Correct				All Correct			

5) How well did the **Aviation Training Tutorial** prepare you to answer all of these knowledge assessment questions (that is, all three sets of questions)?

1	2	3	4	5	6	7
VERY POORLY		FAIRLY WELL			VERY WELL	

APPENDIX D:
BIOGRAPHICAL DATA FORM

Participant Number _____

Biographical Data Form

Please complete the following questions. Any information you provide is voluntary and will be kept strictly confidential. A participant number will be assigned to your responses and in no way will your name be associated with the data. The information you provide will be used only for the purposes of this study.

1. Age: _____
2. Gender: _____ M _____ F
3. Year in school: _____ Freshman _____ Sophomore _____ Junior _____ Senior
4. Major: _____
5. Native language (if not English): _____
6. Do you have any prior knowledge of aviation? _____ Yes _____ No

If yes, please describe:

REFERENCES

- Annett, J. (1989). Training of skilled performance. In A. M. Colley & J. R. Beech (Eds.), *Acquisition and performance of cognitive skills* (pp 61-84). New York, NY: John Wiley & Sons.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191-215.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Berardi-Coletta, B., Buyer, L. S., Dominowski, R. L., & Rellinger, E. R. (1995). Metacognition and problem solving: A process-oriented approach. *Journal of Experimental Psychology*, *21*, 205-223.
- 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.
- Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1) (pp. 77-165). Hillsdale, NJ: LEA.
- Brown, K. G. & Ford, J. K. (2002). Using computer technology in training: Building an infrastructure for active learning. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development: State-of-the-art lessons for practice* (pp. 192-233). San Francisco, CA: Jossey-Bass.

- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, R., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145-182.
- Chi, M. T. H., de Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439-477.
- Craik, F. I. M. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294.
- Cuevas, H. M., Fiore, S. M., Bowers, C. A., & Salas, E. (2004). Fostering constructive cognitive and metacognitive activity in computer-based complex task training environments (Invited paper for Special Issue on Human-Computer Interaction). *Computers in Human Behavior*, *20*, 225-241
- Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science*, *30*, 433-464.
- Davidson, J. E., Deuser, R., Sternberg, R. J. (1994). The role of metacognition in problem solving. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 207-226). Cambridge, MA: The MIT Press.
- Davidson, J. E. & Sternberg, R. J. (1998). Smart problem solving: How metacognition helps. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 47-68). Mahwah, NJ: LEA.

- Dorsey, D. W., Campbell, G. E., Foster, L. L., & Miles, D. E. (1999). Assessing knowledge structures: Relations with experience and posttraining performance. *Human Performance, 12*, 31-57.
- Dunlosky, J., Rawson, K. A., & McDonald, S. L. (2002). Influence of practice tests on the accuracy of predicting memory performance for paired associates, sentences, and text material. To appear in T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 68-92). Cambridge, UK: Cambridge University.
- Evans, W. E., III, Jentsch, F., Hitt, J. M., II, Bowers, C. A., & Salas, E. (2001). Mental model assessment: Is there convergence among different methods? *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 293-296. Santa Monica, CA: HFES.
- Everson, H. T. & Tobias, S. (1998). The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science, 26*, 65-79.
- Fiore, S. M., Cuevas, H. M., & Oser, R. L. (2000). *Supporting knowledge acquisition through enhanced displays: The facilitative effects of diagrams on task performance and mental model development* (Technical Report for the Naval Air Warfare Center Training Systems Division, prepared under Contract No. N61399-99-C-0106). Orlando, FL: University of Central Florida.
- Fiore, S. M., Cuevas, H. M., & Oser, R. L. (2003). A picture is worth a thousand connections: The facilitative effects of diagrams on task performance and mental model development. *Computers in Human Behavior, 19*, 185-199.

- Fiore, S. M., Cuevas, H. M., Scielzo, S., & Salas, E. (2002). Training individuals for distributed teams: Problem solving assessment for distributed mission research. *Computers in Human Behavior, 18*, 729-744.
- Fiore, S. M., Fowlkes, J., Martin-Milham, L., & Oser, R. L. (2000). Convergence or divergence of expert models: On the utility of knowledge structure assessment in training research. *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomic Society, 2*, 427 – 430. Santa Monica, CA: HFES.
- Fiore, S. M., Jentsch, F., Oser, R., & Cannon-Bowers, J. A. (2000). Perceptual and conceptual processing in expert/novice cue pattern recognition. *International Journal of Cognitive Technology, 5 (1)*, 17-26.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906-911.
- Fleishman, E. A. & Mumford, M. D. (1989). Individual attributes and training performance. In I. L. Goldstein (Ed.), *Training and development in organizations* (pp. 183-255). San Francisco, CA: Jossey-Bass.
- 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.
- 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.
- Guilford, J. P. & Zimmerman, W. S. (1981). *Manual of instructions and interpretations for the Guilford-Zimmerman Aptitude Survey* (revised edition). Palo Alto, CA: Consulting Psychological Press.
- Hall, L. K. & Cremer, E. M. (2000, November). Metacognitive judgments of exam performance. Presented at the Forty-first Meeting of the Psychonomic Society, New Orleans, LA. Retrieved from <http://www.owu.edu/~lkhall/psychonomics00.pdf>.
- Hartman, H. J. (2001a). Developing students' metacognitive knowledge and skills. In H. J. Hartman (Ed.), *Metacognition in learning and instruction: Theory, research, and practice* (pp. 33-68). The Netherlands: Kluwer Academic.
- Hartman, H. J. (2001b). Metacognition in science teaching and learning. In H. J. Hartman (Ed.), *Metacognition in learning and instruction: Theory, research, and practice* (pp. 173-201). The Netherlands: Kluwer Academic.
- Hofer, B. K., Yu, S. L. & Pintrich, P. R. (1998). Teaching college students to be self-regulated learners. In D. H. Schunk, & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 57-85). New York, NY: Guilford Publications.

- Jeppesen Sanderson Training Systems. (1996). *Jeppesen Sanderson Private Pilot Exercises Book*. Englewood, CO: Jeppesen Sanderson, Inc.
- Jeppesen Sanderson Training Systems. (1996). *Jeppesen Sanderson Private Pilot Maneuvers Manual (6th ed.)*. Englewood, CO: Jeppesen Sanderson, Inc.
- Jeppesen Sanderson Training Systems. (1996). *Jeppesen Sanderson Private Pilot Manual (15th ed.)*. Englewood, CO: Jeppesen Sanderson, Inc.
- Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: LEA.
- Jonassen, D. H. & Grabowski, B. L. (1993). *Handbook of individual differences, learning, and instruction*. Hillsdale, NJ: LEA.
- Kalyuga, S., Chandler, P., & Sweller, P. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.
- Kelemen, W. L., Frost, P. J., & Weaver, C. A., III. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition, 28*, 92-107.
- King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist, 27*, 111-126.
- Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.

- 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.
- Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 609-616.
- Maki, R. H., Jonas, D., & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychonomic Bulletin & Review*, *1*, 126-129.
- Maqsud, M. (1998). Effects of metacognitive instruction on mathematics achievement and attitude towards mathematics of low mathematics achievers. *Educational Research*, *40*, 237-243.
- Marcus, N. Cooper, M. & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, *88*, 49-63.
- Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, *26*, 46-63.
- Mayer, R. E. (1999). Instructional technology. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay, & M. T. H. Chi (Eds.), *Handbook of applied cognition* (pp. 551-569). Chichester, England: John Wiley & Sons.
- Mayer, R. E. & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, *82*, 715-726.

- Mayer, R. E. & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389-401.
- McInerney, V., McInerney, D. M., & Walsh, H. W. (1997). Effects of metacognitive strategy training within a cooperative group learning context on computer achievement and anxiety: An aptitude-treatment interaction study. *Journal of Educational Psychology, 89*, 686-695.
- Moore, D., Zabrocky, K., & Commander, N. E. (1997). Metacomprehension and comprehension performance in younger and older adults. *Educational Gerontology, 23*, 467-475.
- Oser, R. L., Cannon-Bowers, J. A., Salas, E., & Dwyer, D. J. (1999). Enhancing human performance in technology-rich environments: Guidelines for scenario-based training. In E. Salas (Ed.) *Human/technology interaction in complex systems, Vol. 9* (175-202). Stamford, CT: JAI Press.
- Osman, M. E., & Hannafin, M. J. (1992). Metacognition research and theory: Analysis and implications for instructional design. *Educational Technology Research & Development, 40*, 83-99.
- Paas, F. G. W. C & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*, 737-743.
- Paas, F. G. W. C, Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*, 419-430.

- Pintrich, P. R., Smith, D. A., Garcia, T., McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational & Psychological Measurement, 53*, 801-813.
- Proctor, R. W. & Dutta, A (1995). *Skill acquisition and human performance*. Thousand Oaks, CA: Sage Publications.
- Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*, 181-221.
- 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. (2002). Emerging themes in distance learning research and practice: Some food for thought. *International Journal of Management Reviews, 4*, 135-153.
- 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.
- Schmidt, A. M. & Ford, J. K. (2001, April). Promoting active learning through metacognitive instruction. Paper presented at the 16th Annual Conference of the Society of Industrial and Organizational Psychology, San Diego, CA.

- Schvaneveldt, R. W. (Ed.) (1990). *Pathfinder associative networks: Studies in knowledge organization*. Westport, CT: Ablex Publishing.
- Scielzo, S, Fiore, S. M., Cuevas, H. M, & Salas, E. (2004). Diagnosticity of mental models in cognitive and metacognitive processes: Implications for synthetic task environment training. In S.G. Schiflett, L. R. Elliott, E. Salas, & M. D. Coovert (Eds.), *Scaled worlds: Development, validation, and applications* (pp. 181-199). Aldershot, UK: Ashgate.
- Shanteau, J. (1992). How much information does an expert use? Is it relevant? *Acta Psychologica, 81*, 75-86.
- Smith, E. M., Ford, J. K., & Kozlowski, S. W. J. (1997). Building adaptive expertise: Implications for training design strategies. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 89-118). Washington, D. C.: American Psychological Association.
- Smith-Jentsch, K. A., Ricci, K. E., Campbell, G. E., & Zeisig R. L. (1997). Shaping mental models of the scenario-based training process: A preliminary validation of shipboard instructor training. *Proceedings of the 19th Annual Interservice/Industry Training, Simulation, and Education Conference*. Washington, DC: National Training Systems Association.
- Snow, R. E. (1997). Individual differences. In R. D. Tennyson, F. Schott, N. M. Seel, & S. Dijkstra (Eds.), *Instructional design: International perspectives, Volume 1: Theory, research, and models* (pp. 215-241). Mahwah, NJ: LEA.

- Sorensen, A. G., Brown, K. G., Werner, M. M., & Huntley, J. (2001, April). How do learning differences arise across media? Poster presented at the 16th Annual Conference of the Society of Industrial and Organizational Psychology, San Diego, CA.
- Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science*, *26*, 127-140.
- Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction*, *4*, 295-312.
- Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.
- Sweller, J. & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*, 185-233.
- Van Oostendorp, H. & Goldman, S. R. (Eds.). (1999). *The construction of mental representations during reading*. Mahwah, NJ: LEA.
- Volet, S. E. (1991). Modelling and coaching of relevant metacognitive strategies for enhancing university students' learning. *Learning & Instruction*, *1*, 319-336.
- Weinstein, C. E. & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp. 315-327). New York: Macmillan.
- Winne, P. H. & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Mahwah, NJ: LEA.
- Winne, P. H. & Stockley, D. B. (1998). Computing technologies as sites for developing self-regulated learning. In D. H. Schunk, & B. J. Zimmerman (Eds.), *Self-regulated learning:*

From teaching to self-reflective practice (pp. 106-136). New York, NY: Guilford Publications.

Zimmerman, B. J. (1998). Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models. In D. H. Schunk, & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 1-19). New York, NY: Guilford Publications.