Examining the Impact of Training and Feedback on Visualization-Supported Decision Making under Uncertainty

Jihye Song, Olivia B. Newton, Stephen M. Fiore, Jonathan Coad, Jared Clark,
Corey Pittman, & Joseph J. LaViola Jr.
University of Central Florida

Empirical evaluations of uncertainty visualizations often employ complex experimental tasks to ensure ecological validity. However, if training for such tasks is not sufficient for naïve participants, differences in performance could be due to the visualizations *or* to differences in task comprehension, making interpretation of findings problematic. Research has begun to assess how training is related to performance on decision-making tasks using uncertainty visualizations. This study continues this line of research by investigating how training, in general, and feedback, in particular, affect performance on a simulated resource allocation task. Additionally, we examined how this alters metacognition and workload to produce differences in cognitive efficiency. Our results suggest that, on a complex decision-making task, training plays a critical role in performance with respect to accuracy, subjective workload, and cognitive efficiency. This study has implications for improving research on complex decision making, and for designing more efficacious training interventions to assess uncertainty visualizations.

INTRODUCTION

Decision making under uncertainty is a ubiquitous issue for operations in complex environments. The use of visualizations as decision aids is an increasingly common research topic in a variety of complex task domains (Kinkeldey, MacEachren, Riveiro, & Schiewe, 2017). To this end, studies of uncertainty visualization have attempted to formulate theoretically grounded design guidelines for visualization to support decision-making processes (e.g., Meyer et al., 2010; MacEachren et al., 2012; Kinkeldey, MacEachren, Riveiro, & Schiewe, 2017). The empirical evaluation of uncertainty visualization for this effort necessitates the use of complex experimental tasks. Specifically, to ensure the ecological validity and generalizability of this line of research, experimental tasks must be sufficiently representative of those carried out by experts in their relevant task domain. The generalizability of this body of work is also limited by the fact that experiments are often conducted with naïve participants who do not comprehend complex experimental tasks as well as experts. For example, study results showing a lack of an effect of an uncertainty visualization could be attributed to an ineffective visualization *or* to an inadequate understanding of the task by participants. Taken together, these factors impede the ability to delineate the relationship between uncertainty visualizations and decision making. As such, it is critical that training for complex tasks is adequate and effective for the sampled population. Otherwise, inferences about the effects of uncertainty visualization interventions are problematic.

We seek to build on recent work investigating training for decision-making tasks utilizing uncertainty visualizations (Fiore et al., 2018). We argue that in order to accurately evaluate uncertainty visualizations as a decision aid, greater attention must be given to variables that could influence performance, such as degree of task comprehension, which may be influenced by training prior to experimental tasks.

Testing is often used to examine the efficacy of training by assessing participants' knowledge acquisition. Additionally, testing itself can influence knowledge acquisition, notably improving long-term information retention (Roediger & Karpicke, 2006; Karpicke & Roediger, 2008). Related to the effect of testing on learning, feedback has been researched extensively. Providing feedback to test takers can facilitate learning and long-term retention (e.g., Butler, Karpicke, & Roediger, 2008; Van Buskirk, 2011). However, although there is significant evidence that feedback in general may be an effective tool, there is less research focusing on how specific features of feedback affect learning and performance.

In the context of training, testing combined with feedback can facilitate learning by providing important information about performance to guide future performance. The extant literature on feedback delivery offers several conflicting perspectives on the mechanisms influencing feedback effectiveness. For example, with respect to the timing of feedback, while some researchers have found immediate feedback to be superior to delayed (Van Buskirk, 2011), an alternative argument presents delayed feedback as a superior learning aid because it reduces memory load (e.g., Brackbill & Kappy, 1962). Proponents of delayed feedback argue that reduced demand on memory results in more efficient encoding and improved retention of essential information (Butler et al., 2007; 2008; Smith & Kimball, 2010; Soderstrom et al., 2016).

We also seek to improve upon how learning and performance are measured by using more sophisticated assessments. Our goal is to converge on a richer understanding on how interventions alter cognitive processes. In particular, we examine how metacognition is altered and how cognitive efficiency, a measure combining workload with performance, is changed by training. Metacognition is a multidimensional phenomenon involving knowledge of one's cognitions and regulation of those cognitions (Schraw, 1998). A considerable body of literature has explored the relationship between metacognitive processes and learning outcomes in a variety of domains (e.g., Gourgey, 1998; Mayer, 1998; Sternberg, 1998). For example, Kruger and Dunning (1999) found that errors in metacognition were related to academic performance such that those poorly calibrated in predicting their performance also performed poorly. Others show that high performers tended to be accu-

 DOI 10.1177/1541931218621329

rate when predicting performance, but low performers were less accurate in predictions (Hacker et al., 2000). Further, the lowest performers were consistently poor at predictions and postdictions (see also Maki et al. 2005). We set out to determine how learning during complex task training alters metacognitive biases in prediction and postdiction of performance.

We additionally emphasize how cognitive efficiency, evolving out of instructional efficiency, can better diagnose training effectiveness. This assesses the relationship between subjective assessment of workload and overall task performance (Fiore et al., 2006; Paas & Van Merrienboer, 1993). By standardizing measures of performance and workload and computing difference between scores, a more diagnostic picture of training emerges. Positive scores indicate relative performance is higher than workload. This can be interpreted as showing that some intervention led to more efficient cognitive processing. Negative scores indicate relative performance was less than relative workload. Prior research finds that decision support augmented with graphical displays leads to higher cognitive efficiency when compared to those not using such displays (see Fiore et al., 2017; Johnston et al. 2013).

Summary

Visualizations garner significant interest in complex systems as a means of supporting decision making under uncertainty (Kinkeldey et al. 2017; Smith Mason, et al., 2017). Yet, theoretical perspectives of the relationship between uncertainty visualization and complex cognition are scarce (MacEachren, 2015; Sedig & Parson, 2013). Further, empirical studies of uncertainty visualization have yielded mixed findings, limiting the generalizability of visualization studies, thus hindering a coherent assessment of the concept. Fiore et al. (2018) suggest that differences in results across studies may be due, in part, to inadequate or inconsistent training for complex experimental tasks. The results of their experiment demonstrate the importance of adequate training for subsequent performance in a decision-making task. Further, this highlights a limitation of uncertainty visualization studies that lack consistency in the administration of training, given that training type may influence performance.

In light of prior work, we designed the present experiment to examine a novel approach to training that employs distributed feedback during training. As such, we build on our prior work examining how training influences performance on complex visualization tasks associated with decision making under uncertainty. To this end, we assessed the effect of training and feedback on learning (i.e., knowledge acquisition) and performance in a complex task (i.e., knowledge application).

METHODS

This effort builds upon a previous study assessing the value of different types of training for decision making under uncertainty (Fiore et al., 2018). The present study specifically examines the effect of variations in testing and feedback presented during or immediately after training. For the decision-making task, we developed a scenario requiring participants to evaluate the effect of uncertainty on their decisions and thus presented participants with a complex task requiring both acquisition and application of knowledge.

Participants

We recruited 200 participants through Amazon's MTurk (41% female and 59% male, mean age = 32.9 years). Participants were required to identify English as their primary language. Based on previous studies, participants were compensated $2.00 USD for completing the study. To increase motivation, participants also earned a $1.00 bonus if they scored in the top ten percent during the decision-making task.

Experimental Design

We used a between-subjects design to investigate the effect of testing and feedback distribution for a decision-making task presented in the form of a game. The training was partitioned into four sections. In the training, participants reviewed text information and corresponding images, including descriptions of task objectives, relevant capabilities and limitations of game elements, and how uncertainty is represented in the game. Participants were randomly assigned to one of four conditions. Each group received the same training content. Figure 1 depicts an example of an image from the training.

Figure 1. Example of an image from the "Your Vehicles" training module

Feedback Delivery

Feedback distribution served as the independent variable. Feedback took the form of a knowledge acquisition assessment followed by a list of correct answers (described next). Based on where feedback was presented for each condition, participants either completed all four assessment sections together at the end of the training or throughout the breaks between modules. This was done to ensure all feedback sections were presented consistently relative to their respective assessment sections. Thus, for this experiment, feedback included both a knowledge acquisition assessment, as well as presentation of correct answers. Feedback delivery was manipulated at four levels: no feedback (control); massed feedback after training; distributed feedback at two points (halfway through the training and after the training); and distributed feedback after each training module.

Knowledge Acquisition and Application

During the training, learning was assessed using tests of knowledge acquisition. Following the training, all participants completed a complex decision-making task requiring the use of visualizations of uncertainty in a set of simulation vignettes.

Knowledge acquisition. This assessment comprised four sections of multiple-choice questions corresponding to the four training modules. Within these, two types of questions were presented: (1) Recognition, which required participants to identify game components based on images, and (2) Declarative, which required knowledge of concepts covered in the training. This assessment was used to gauge level of comprehension during the training, as well as provide feedback.

Knowledge application. To test participants' ability to apply and integrate their knowledge of training, we developed a decision-making task in the form of simulation vignettes. These consisted of scenarios designed as an analogue to a Naval Intelligence Unit drug interdiction task. The scenario, called "The Party Game," required participants to recover party supplies that were lost at sea in order to throw a party on a tropical island. Participants maximized points by collecting a variety of supplies using minimal resources in the least amount of time. Adding to the complexity, participants had to consider the capabilities of different vehicles, as well as uncertainty regarding position of supplies or rivals, presence of rivals attempting to steal supplies, and weather conditions. Uncertainty was depicted using spaghetti plots and participants were instructed to use these visualizations to help them select the best course of action. The task was presented through the use of images from the simulation testbed (i.e., simulation vignettes). The vignettes differed with respect to the weather, uncertainty, supplies, vehicles, and rivals in the environment.

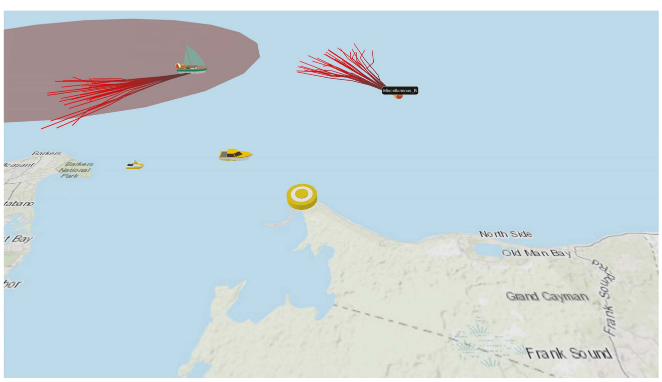

Figure 2. Example of a simulation vignette.

Workload

Each multiple-choice question in the Knowledge Acquisition and Knowledge Application assessments was followed by a subjective workload assessment. For this workload assessment, participants used a 7-point Likert-type scale ranging from 1 (*very easy*) to 7 (*very difficult*) to rate the difficulty of the preceding assessment item.

Cognitive Efficiency

We measured participants' cognitive efficiency (CE) by examining the relationship between workload and performance accuracy. CE is derived by combining standardized workload scores (z_w) with standardized performance scores (z_p). As described in Fiore et al. (2006), such scores can be represented as the perpendicular distance from a line representing a level of zero efficiency (see Equation 1).

$$CE = \frac{(z_p - z_w)}{\sqrt{2}} \tag{1}$$

Because these are standardized scores, this results in positive and negative values that hover around a mean of 0. Positive scores indicate CE in that there is relatively better performance in proportion to reported workload, whereas negative scores indicate cognitive inefficiency (i.e., relative performance is less than relative workload).

Metacognitive Biases

Our goal was to understand the degree to which metacognitive accuracy may vary as a function of learning a complex task (Cuevas et al., 2004). To determine accuracy, participants made subjective assessments of performance. First, immediately following completion of the training, participants were asked to judge how many correct answers they would get on questions on the material just presented. Second, following completion of the knowledge assessment task, participants were asked to report how well they thought they did on the test overall. The difference between performance and predication (before test) and postdiction (following test) is an indication of metacognitive bias (calculated as "Prediction – Performance"). Scores closer to zero indicate relatively accurate metacognition in that the trainee is able to accurately gauge how well they understood the material and positive and negative scores indicate over- or under-predicting performance.

Research Hypotheses

Hypothesis 1: Performance accuracy. Our main question is whether participants will benefit from training on uncertainty visualization decision making. That is, regardless of feedback condition, participants documenting greater learning during the training will perform better on the knowledge application assessment (uncertainty visualization vignettes). Next, we predict that participants in the Individual Distributed Feedback condition will outperform participants in the Grouped Distributed Feedback, Massed Feedback, and No Feedback conditions. Specifically, there will be a significant effect of the feedback condition on the knowledge application assessment.

Hypothesis 2: Workload. Our main question is whether workload is affected by training on uncertainty visualization decision making. That is, regardless of feedback condition, participants documenting greater learning during the training will show lower workload on the knowledge application assessment (uncertainty visualization vignettes). Next, we predict that participants in the Individual Distributed Feedback condition will report lower workload than the Grouped Distributed Feedback, Massed Feedback, and No Feedback conditions. Specifically, there will be a significant effect of the feedback condition on overall workload experienced during knowledge assessment.

Hypothesis 3: Cognitive Efficiency. Our main question is whether CE is altered dependent upon learning effectiveness for uncertainty visualization decision making. That is, regardless of feedback condition, participants documenting greater learning during the training will show greater CE on the knowledge application assessment, that is, on the uncertainty visualization vignettes. Next, we predict that participants in the Individual Distributed Feedback condition will show greater CE than the Grouped Distributed Feedback, Massed Feedback, and No Feedback conditions. Specifically, there will be a significant effect of the feedback condition on overall CE on the knowledge application assessment.

Hypothesis 4: Metacognitive Bias. Our main question is whether metacognitive processes are altered dependent upon learning effectiveness for uncertainty visualization decision making. That is, regardless of feedback condition, participants documenting greater learning during training will show better metacognition as demonstrated by lower bias. Next, we predict that participants in the Individual Distributed Feedback condition will show better metacognition than the Grouped Distributed Feedback, Massed Feedback, and No Feedback conditions. Specifically, there will be a significant effect of the feedback condition on metacognitive bias scores on the knowledge application assessment.

Procedure

Participants were directed to our study hosted by Qualtrics. Upon providing informed consent, participants were randomly assigned to one of the four conditions. Participants received the same training and assessment questions in the order determined by condition. Following training, participants completed the knowledge acquisition assessment. Following this, for all conditions except No Feedback, participants were given feedback consisting of the correct answers to the questions. Next, participants completed the knowledge application assessment. For each of the simulation vignette questions, three of the response options were technically possible (i.e., followed the rules of the game), but only one of these was optimal (i.e., maximized points). The fourth response option was completely incorrect, in that it violated the game's rules. Throughout the assessments, participants were asked to rate how difficult they found each item. Finally, participants filled out a demographic questionnaire.

RESULTS

There were no effects of feedback distribution on the knowledge application (simulation vignette) assessment. Therefore, we only report the effects on our primary hypotheses about how learning effectiveness from training, influences performance on a complex decision-making task (our simulation vignettes), as well as on workload, CE, and metacognition. For these analyses, amount of learning was determined by analyzing performance on the knowledge acquisition assessment for the training modules. Based upon this, a median split was used to divide participants into two groups (Low Learners versus High Learners). With this breakdown we test how learning during training on a complex decision-making task can influence performance accuracy, workload, and CE experienced during the simulation vignettes utilizing uncertainty visualizations.

First, there was a significant effect of learning on performance accuracy during the uncertainty visualization simulation vignettes, $F(1, 198) = 33.3$, $p < .001$, $\eta_p^2 = .14$, observed power = 1.0. Accuracy was lower for Low Knowledge Acquisition participants ($M = .28$, $SD = .15$) compared to High Knowledge Acquisition participants ($M = .41$, $SD = .18$).

Second, there was a significant effect of learning on reported workload during the simulation vignettes, $F(1, 198) = 14.2$, $p < .01$, $\eta_p^2 = .04$, observed power = .78. Reported workload was greater for Low Knowledge Acquisition participants ($M = 4.7$, $SD = 1.3$) compared to High Knowledge Acquisition participants ($M = 4.4$, $SD = 1.4$).

Third, there was a significant effect of learning on CE for the simulation vignettes, $F(1, 198) = 32.2$, $p < .001$, $\eta_p^2 = .14$, observed power = 1.0. CE was lower for Low Knowledge Acquisition participants ($M = -.41$, $SD = .88$) compared to High Knowledge Acquisition participants ($M = .39$, $SD = 1.1$).

Fourth, there was a significant effect of learning on metacognitive prediction scores for the simulation vignettes, $F(1, 198) = 9.72$, $p < .01$, $\eta_p^2 = .05$, observed power = .874. For prediction of performance, Low Knowledge Acquisition participants ($M = .376$, $SD = .324$) showed greater metacognitive bias compared to High Knowledge Acquisition participants ($M = .249$, $SD = .248$). Relatedly, there was a significant effect of learning on metacognitive postdiction scores for the simulation vignettes, $F(1, 198) = 17.9$, $p < .001$, $\eta_p^2 = .08$, observed power = .98. For postdiction of performance, Low Knowledge Acquisition participants ($M = .308$, $SD = .312$) showed greater metacognitive bias compared to High Knowledge Acquisition participants ($M = .144$, $SD = .286$).

DISCUSSION

This study set out to determine how training alters acquisition and application of knowledge for a complex decision-making task associated with visualization of uncertainty. First, we studied how comprehension of training content alters performance on decision-making tasks. Second, we examined how varieties of feedback, during training, alters performance. Across these, we examined the effects on workload, metacognition, and cognitive efficiency, a combinatory metric of workload and performance.

The results show that acquisition of task related content during training alters performance when using visualizations to make decisions under uncertainty. This is in line with prior work showing how training relates to performance on tasks using uncertainty visualizations (Fiore et al., 2018). We find, though, that feedback did not alter performance. Distributing testing and feedback within training did not have a significant effect on accuracy, workload, or cognitive efficiency.

Although our results suggest distributing feedback does not have a significant effect on learning and performance, this could be due to the fact that our training and assessments were not long enough for feedback distribution to make a significant difference. It is possible that with a longer training and assessment requiring greater knowledge acquisition, distributing testing and feedback in smaller chunks across training would reduce cognitive load to a greater extent than with a relatively short training and assessment. Nonetheless, these findings illustrate the need to closely examine variables that could influence learning in order to gain a better understanding of participants' task comprehension when assessing uncertainty visualizations. Another potential limitation is that, for conditions where feedback was presented, participants were only shown the correct responses; they were not explicitly told which or how many questions they had answered correctly. Therefore, it is possible that if participants did not recall their own responses, feedback might not have provided them with enough information to assess their performance.

CONCLUSION

The present study has both theoretical and practical implications. By critically examining factors that may influence learning, we can better understand learning processes in general, as well as in specific contexts such as during computer-based training. In terms of practical implications, the present research can be applied to future work investigating the effectiveness of uncertainty visualizations. By understanding how training and testing influence learning and task performance, we can more easily interpret the effect of uncertainty visualizations on decision making. In brief, uncertainty visualization research must ensure that training for complex tasks is effective for naïve populations. In the absence of this, inferences about the effects of uncertainty visualization interventions are problematic.

In sum, this work contributes to the growing body of research on training for complex tasks involving decision making under uncertainty. Future studies should seek to determine whether our findings are replicable both for shorter and longer tasks and extend them to different contexts with complex uncertainty visualization tasks to improve the generalizability of our results. For example, future work can examine differences in learning and performance across naïve, novice, and expert populations to improve the design of training for complex tasks. Additionally, future work can explore the influence of individual differences on learning and decision making in order to identify the most appropriate training methods for specific subpopulations.

ACKNOWLEDGEMENT

This work was supported by the Office of Naval Research grant N00014-15-1-2708, under the Command Decision Making program. The views and opinions contained in this article are the authors and should not be construed as official or as reflecting the views of the University of Central Florida or the Office of Naval Research.

REFERENCES

Brackbill, Y., & Kappy, M. S. (1962). Delay of reinforcement and retention. *Journal of Comparative and Physiological Psychology*, *55*(1), 14.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13*(4), 273.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 918.

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.

Fiore, S. M., Song, J., Newton, O. B., Pittman, C., Warta, S. F., LaViola, J. J. (2018). Determining the effect of training on uncertainty visualization evaluations. *Proceedings of the 9th Annual Conference of Applied Human Factors and Ergonomics*. Berlin: Springer.

Fiore, S. M., Scielzo, S., Jentsch, F., & Howard, M. (2006). Effects of discrimination task training on X-ray screening decision. *Proceedings of the 50th Annual Meeting of the HFES* (pp. 2610-2614). Santa Monica, CA: Human Factors and Ergonomics Society.

Fiore, S. M., Warta, S. F., Best, A., Newton, O., & LaViola, J. J. (2017). Developing a theoretical framework of task complexity for research on visualization in support of decision making under uncertainty. *Proceedings of the Annual Meeting of the HFES, 61*(1), 1193–1197.

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

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92,* 160–170.

Johnston, J., Fiore, S. M., Paris, C., & Smith, C. A. P. (2013). Application of Cognitive Load Theory to developing a measure of team cognitive efficiency. *Military Psychology, 25*(3), 252-265.

Karpicke, J. D. & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966-968.

Kinkeldey, C., MacEachren, A. M., Riveiro, M., & Schiewe, J. (2017). Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science, 44*(1), 1–21.

MacEachren, A. (2015). Visual Analytics and Uncertainty: It's Not About the Data. In E. Bertini & J.C. Roberts, (Eds.), *EuroVis Workshop on Visual Analytics.* Geneva, Switzerland: Eurographics Association.

MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., & Gahegan, M. (2012). Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics, 18*(12), 2496–2505.

Maki, R. H., Shields, M., Wheeler, A., E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology, 97,* 723-731.

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

Meyer, J., Thomas, J., Diehl, S., Fisher, B. D., Keim, D. A., Laidlaw, D., ... & Ynnerman, A. (2010). From visualization to visually enabled reasoning. *Scientific Visualization: Advanced Concepts, 1*, 227-245.

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737-743.

Roediger III, H. L. & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249-255.

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

Sedig, K., & Parsons, P. (2013). Interaction design for complex cognitive activities with visual representations: A pattern-based approach. *AIS Transactions on Human-Computer Interaction, 5*(2), 84-133.

Smith Mason, J., Retchless, D. & Klippel, A. (2017). Domains of uncertainty visualization research: A visual summary approach. *Cartography and Geographic Information Science, 44*(4), 296-309.

Smith, T. A. and Kimball, D. R. (2010). Learning from feedback: spacing and the delay-retention effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 80–95.

Soderstrom, N. C., Kerr, T. K., & Bjork, R. A. (2016). The critical importance of retrieval—and spacing—for learning. *Psychological Science, 27*(2), 223–230.

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

Van Buskirk, W. L. (2011). *Investigating the optimal presentation of feedback in simulation-based training: An application of the cognitive theory of multimedia learning* (Unpublished doctoral dissertation). University of Central Florida, Orlando, Florida.