2010

# Adaptive Feedback In Simulation-based Training

Deborah Billings
*University of Central Florida*

# ADAPTIVE FEEDBACK IN SIMULATION-BASED TRAINING

by

DEBORAH RAE BILLINGS
B.A. Agnes Scott College, 2002
M.S. University of Central Florida, 2008

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 Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2010

Major Professor: Richard D. Gilson

## ABSTRACT

Feedback is essential to guide performance in simulation-based training (SBT) and to refine learning. Generally outcomes improve when feedback is delivered with personalized tutoring that tailors specific guidance and adapts feedback to the learner in a one-to-on environment. Therefore, emulating by automation these adaptive aspects of human tutors in SBT systems should be an effective way to train individuals.

This study investigates the efficacy of automating different types of feedback in a SBT system. These include adaptive bottom-up feedback (i.e., detailed feedback, changing to general as proficiency develops) and adaptive top-down feedback (i.e., general feedback, changing to detailed if performance fails to improve). Other types of non-adaptive feedback were included for performance comparisons as well as to examine the overall cognitive load.

To test hypotheses, 130 participants were randomly assigned to five conditions. Two feedback conditions employed adaptive approaches (bottom-up and top-down), two used non-adaptive approaches (constant detailed and constant general), and one functioned as a control group (i.e., only a performance score was given). After preliminary training on the simulator system, participants completed four simulated search and rescue missions (three training missions and one transfer mission). After each training mission, all participants received feedback relative to the condition they were assigned.

Overall performance on missions, knowledge post-test scores, and subjective cognitive load were measured and analyzed to determine the effectiveness of the type of feedback. Results indicate that: (1) feedback generally improves performance, confirming prior research; (2) performance for the two adaptive approaches (bottom-up vs. top-down did not differ significantly at the end of training, but the bottom-up group achieved higher performance levels significantly sooner;  (3) performance for the bottom-up and constant detailed groups did not differ significantly, although the trend suggests that adaptive bottom-up feedback may yield significant results in further studies. Overall, these results have implications for the implementation of feedback in SBT and beyond for other computer-based training systems.

This effort was inspired by Dad and is dedicated to him in loving memory.

This work is also dedicated to Mom, who is an amazing lady.

## ACKNOWLEDGMENTS

This dissertation was funded by the U.S. Army Research Institute through the Consortium Research Fellows Program, partnered with the University of Central Florida. Positions expressed in this dissertation are those of the author and do not represent an official position of the U.S. Army, the Army Research Institute, or the Consortium of Universities. I wish to extend special gratitude to Dr. Stephen Goldberg and Dr. Robert Ruskin who supported this research project and gave me the opportunity to succeed.

I am also sincerely grateful to Dr. Richard Gilson, my advisor and committee chair, for preparing me for this endeavor, giving me timely feedback, supporting me throughout my academic career, and constantly challenging me to go above and beyond expectations. Without his dedication and confidence in my abilities, I would not have made it this far. I also wish to thank Dr. Mustapha Mouloua and Dr. Valerie Sims for their insightful comments, guidance, and assistance throughout the project. In addition, I would like to express deep appreciation to Dr. Paula Durlach and Dr. Heather Priest, both of whom have been wonderful mentors and friends to me. Without their willingness to share statistical expertise, research knowledge, and advice, I would not have been able to complete this project so quickly.

Finally, I would like to thank my personal friends and family for their love and support throughout this process. For my doctoral cohort Amanda Surprenant, Jing Jing Wang-Costello, and Stephanie Quinn: I was fortunate enough to have your friendship, support, and encouragement during the extent of the doctoral program. For John, I was blessed to have you walk into my life when I needed you most. Lastly, but certainly not least, for my mom (my editor), my aunt Raynette, my grandparents, and the rest of my family: thank you for believing beyond a shadow of a doubt that I could do it.

## TABLE OF CONTENTS

## LIST OF FIGURES

**LIST OF TABLES**

## LIST OF ACRONYMS

| | |
|---|---|
| ANCOVA | Analysis of Covariance |
| CLT | Cognitive Load Theory |
| CLQ | Cognitive Load Questionnaire |
| CV | Covariate |
| DoD | Department of Defense |
| ERE | Expertise Reversal Effect |
| FiST | Fire Support Team |
| FO | Forward Observer |
| GEM | Game Experience Measure |
| ITS(s) | Intelligent Tutoring System(s) |
| LO | Learning Objective |
| LTM | Long Term Memory |
| MHQ | Motion History Questionnaire |
| MOUT | Military Operations in Urban Terrain |
| ONR | U.S. Office of Naval Research |
| SBT | Simulation-based training |
| TAFS | Texting and Automated Feedback System |
| VGE | Video Game Experience |
| WM | Working Memory |
| ZPD | Zone of Proximal Development |

# CHAPTER ONE: INTRODUCTION

## Importance of Simulation-Based Training

Simulation-based training (SBT) systems offer a promising method of training because they provide realistic, versatile environments where individuals can learn new information, directly apply this information to simulated tasks, and master complex material (Menaker, Coleman, Collins, & Murawski, 2006). SBT has an advantage over conventional classroom teaching techniques in that it takes people out of a passive learning environment and allows them to gain experience at directly applying the information in a simulated environment that mimics the real-world (Nicholson, Fiore, Vogel-Walcutt, & Schatz, 2009). While it is a good training tool for these reasons, SBT has the potential to become an enormously successful training tool that can cater to the individual needs of students, if designers can understand how to integrate the appropriate pedagogical techniques. When instructional components are designed and implemented properly, individuals are better able to learn, assimilate, and apply information (even if they are required to perform in demanding environments; Nicholson et al., 2009).

Research has demonstrated that SBT provides an effective alternative to textbook or classroom learning (Tichon, 2007) because it can support learning and help learners create and maintain mental models for new information (Cuevas, Fiore, Bowers, & Salas, 2004). Consequently, SBT systems are used in many diverse fields and domains, including many practical applications in military domains (Chang, 2009). Among the advantages, SBT systems are very accessible and can provide an authentic simulated environment (where the curriculum is the scenario itself) in which Soldiers can learn required skills (tactical skills, marksmanship, cognitive training, procedural skills, etc.; Chang, 2009). Second, SBT gives Soldiers a cognitive benefit; they are given the ability to visualize and practice their actions in a simulation before a live exercise, which can give them an added advantage so that they can experience the exercise before doing a live version (i.e., they can recall what they learned/saw in the simulated exercise; Waldman, 2009). Third, SBT systems have the advantage of being cheaper and safer than training Soldiers using real equipment, ammunition, vehicles, etc (Pine, 2009). Fourth, Soldiers can go through many different scenarios and get extremely varied experience in a short time (which is one of the biggest draws of SBT); with live exercises, much more time, manpower, and resources would be needed to provide as many varied scenarios (Pine, 2009). Finally, scenarios in SBT can be replayed as many times as necessary, and Soldiers are able to practice procedures for malfunctions or systems failures without putting their own lives at risk (Pine, 2009). While nothing is as good as the real thing, these training simulations can help Soldiers build confidence for performing in the real live environment when that environment is not readily available for them to train in (Chang, 2009).

In addition to military applications, SBT systems are used in the medical arena to teach medics how to act quickly and effectively, in law enforcement to train procedures, and in emergency organizations to train personnel how to manage natural and man-made disasters (Salas & Cannon-Bowers, 2001). The outlined advantages of SBT for Soldiers (e.g., cognitive benefits, cost benefits, etc.) also apply to these domains. Many researchers think that SBT will only become more prevalent in a variety of domains in the future (Waldman, 2009). For this reason, research needs to examine ways to incorporate pedagogical interventions so that the instructional effectiveness of these systems can be optimized.

*Role of Feedback*

The ultimate goal of SBT is to help individuals develop cognitive and decision-making skills for performing (and mastering) more complex tasks, especially when there is uncertainty in environments (for example, training airline pilots, air-traffic controllers, Soldiers, and medical doctors; Tichon, 2007). Often, instructors will rely on trainee exposure to a simulation or simulated training task in hopes that this will promote learning (Ward, Williams, & Hancock, 2006). However, simply playing a serious video game or interacting with a simulation alone does not automatically lead to a meaningful learning experience. For example, individuals may not always understand how to interact with the simulation, and they may not realize when or if they perform an action incorrectly (Johnson & Rickel, 1996). Instead, a simulation should serve as a vessel to deliver instruction and implement pedagogical principles to ensure learning takes place (Ward et al., 2006). In order to confirm learning occurs and that a SBT system is instructionally effective, student learning must be constantly evaluated. This evaluation makes sure the training does what it is intended to do. For example, if unmanned vehicle operators train via a computer-based simulated exercise and then amass a significant number of failed missions in the field, then the simulation certainly does not provide adequate training to the operators. One reason such a failure can occur in simulations is due to insufficient, poorly designed feedback during training. To guarantee learning takes place, training should support the processes at the core of learning: selecting (focusing on the relevant information), organizing (forming mental representations of the information), and integrating (combining the new information with previous knowledge existing in long term memory; Clark & Mayer, 2008). Feedback interventions should also support individuals in these ways.

No standard guidelines exist for implementing feedback in SBT, yet feedback is very important because it allows a trainee to compare discrepancies between his or her performance and the required or goal performance for the task at hand (Kluger & DeNisi, 1996). In this way, individuals are able to reflect, deal with the discrepancies, actively learn from their errors, and improve task performance and retention. Feedback has the potential to significantly improve learning and performance outcomes; however, there is continuing discussion about how and when to deliver feedback (Shute, 2008; Mason & Bruning, 2001; McLaughlin, Rogers, & Fisk, 2008). Narciss (2008) notes that, "modern information technologies increase the range of feedback strategies that can be implemented in computer-based learning environments; however, the design and implementation of feedback strategies are very complex tasks that are often based more on intuition than on psychologically sound design principles" (p. 126). Consequently, research must be conducted to empirically determine the most appropriate ways to use technology to administer feedback in SBT environments, which may not always align with strategies that are thought to be "intuitive."

Feedback is an area that has been researched extensively over the years. Unfortunately, the inconsistent use of terminology makes interpreting this research difficult. The table found in Appendix A outlines feedback terms used in the current research and other terms that are commonly used interchangeably or to describe very similar types of feedback in the literature.

*Future Directions of Simulation-Based Training (SBT)*

In terms of feedback, the future of SBT is likely to involve some form of adaptive, individualized support. Adaptive instruction dynamically changes in response to the learning needs, personal abilities, skills, and other individual differences of each individual student (Lee & Park, 2008, Shute & Zapata-Rivera, 2007, Mills & Ragan, 1994). The rationale is that every person possesses different levels of prior knowledge, skills, and abilities and hence may need varying degrees of support and flexibility during training. Flexibility to meet individual needs could be extremely beneficial for training systems, as demonstrated by Bloom's (1984) research where students performed significantly better with flexible one-to-one tutoring than those who received classroom instruction. Adaptive systems may eventually be capable of capturing information about the trainees as they perform the task, analyzing the current state of the trainees, selecting the appropriate intervention, and presenting that information to the trainees (Shute & Zapata-Rivera, 2007).

Several computer-based training systems exist that include some adaptive aspects of instruction; these are called Intelligent Tutoring Systems (ITSs). ITSs use artificial intelligence to automatically provide intelligent, personalized instruction to trainees, monitor their progress throughout the task, diagnose errors, and deliver suitable feedback without the need for a human tutor to be present during the learning process (Schatz, Bowers, & Nicholson, 2009; Park & Lee, 2004). The theoretical basis for ITSs comes from Bloom's (1984) work on how individualized instruction optimizes learning. The ultimate goal of an ITS is to mirror how a human tutor interacts and adapts to a student during one-on-one instruction, which is the most effective yet expensive way to teach (Corbett & Anderson, 2001). Researchers have been studying how to create effective ITSs for over 40 years, but the goal of successfully imitating all aspects of a human tutor in an ITS remains elusive (Kenny & Pahl, 2009).

ITSs have been somewhat successful, but they do not facilitate experiential learning the same way that SBT systems do. Most existing ITSs teach very well-established domains which follow rules and have objectively correct procedures and answers, such as algebra, geometry, physics, and biology principles. According to Nicholson et al. (2009), this is one of the disadvantages of an ITS; although an ITS can be efficient in terms of training straightforward declarative information, it does not give trainees a dynamic and applied experience in the same way that a simulation does. Consequently, hybrid training systems called situated tutors have evolved that attempt to mesh ITS with SBT. These situated tutors incorporate the benefits of the intelligent components of adaptive instruction and the applied context of SBT (Nicholson et al., 2009). While empirical support for the effectiveness of these types of systems is growing (Mangos & Johnston, 2009), in the literature it can be hard to discriminate the pedagogical strategies that may be optimally effective for a situated tutor (or a SBT system) as opposed to an ITS that simply trains static, declarative information such as mathematics or physics.

Adaptive instruction in SBT systems and ITSs can be approached in three different ways: from a macro-level approach, from an aptitude-treatment interaction approach, and from a micro-level approach (Park & Lee, 2004). On the macro-level, pre-task measures (e.g. cognitive ability and instructional goals) are taken, and appropriate instructional components are selected and adapted to that individual based on those pre-measures before instruction even begins (Park & Lee, 2004). For example, sequencing task difficulty and dynamic problem selection are considered macro-level approaches (Camp, Paas, Rikers, & van Merriënboer, 2001). Here, training is tailored to pre-existing traits, abilities, and limitations of an individual with the goal of enhancing learning. To implement adaptive instruction using an aptitude-treatment interaction approach, instructional strategies are adapted to specific student characteristics or aptitudes that are measured before instruction begins. The goal is to select certain instructional strategies that best help learners with particular aptitudes (Park & Lee, 2004). For example, individuals with a particular learning style may learn best using certain strategies, while those with a different learning style may learn best using different instructional strategies. Finally, on the micro-level, adaptive instruction involves on-going assessment measures to continuously diagnose performance, knowledge level, or state characteristics throughout the learning process (Park & Lee, 2004). An individual's unique learning needs change throughout instruction, and this micro-level approach provides tailored instruction for those changing needs (i.e., instructional strategies change as an individual's performance or attitudes change; Park & Lee, 2004). A combination of these approaches is sometimes used, although micro-adaptive instruction is more likely to be sensitive to students' needs because it assesses learning needs during instruction (Park & Lee, 2004). Therefore, the current research focuses on how to implement adaptive feedback in SBT systems at a micro-level, based on knowledge levels measured via ongoing performance assessments.

## Purpose of Current Research

Decreased training time and increased learning are two main aims of training programs, and both of these things can help reduce training costs over time (Clark, Nguyen, & Sweller, 2006). Adaptive training and SBT systems may collectively help to reach those goals, and researchers agree that infusing adaptive training components into training systems are important to future success of the systems (Shute & Zapata-Rivera, 2007; MYMIC LLC, 2004). Technology is becoming increasingly sophisticated, which gives researchers and designers the tools to develop and implement extremely efficient micro-adaptive instructional techniques (Park & Lee, 2004). Thus, the tools are available, but how can our current knowledge of cognitive and pedagogical principles be combined to create SBT systems that utilize adaptive feedback effectively at the micro-adaptive level?

While technology enables the inclusion of training features that were once tricky (or impossible) to incorporate, training components are frequently included in SBT systems and ITSs that are "…recommended purely because they are now possible rather than because there [is] evidence for their cognitive effectiveness or even desirability" (Sweller, 2008, p. 380). Designers should not simply add adaptive feedback or other instructional components into SBT systems without solid proof of their usefulness. Empirical evidence is needed to support the addition of training components. It is not acceptable to include components in a system simply because modern technology allows it. In particular, the design of adaptive feedback components needs to have a sound theoretical rationale and be empirically tested to determine how specific feedback interventions will impact learning and performance throughout training (Mangos & Johnston, 2009).

The goal of the current research is to investigate the efficacy of a theoretically-based method for administering adaptive feedback in SBT. This feedback research is integral to the future development of SBT systems and ITSs. Sweller (2008) warns researchers that the effectiveness of feedback can be traced back to whether or not the instructional designer was guided by theoretical assumptions of human cognition. Consequently, the present research considers the Cognitive Load Theory (CLT) and the Expertise Reversal Effect (ERE) in the design and implementation of adaptive feedback, and testable hypotheses are developed to conduct empirical research on the effects of adaptive feedback in a SBT system.

# CHAPTER TWO: BACKGROUND

## Functions of Feedback

In instructional or training contexts, feedback can be defined as "…post-response information that is provided to learners to inform them of their actual state of learning or performance" (Narciss, 2008, p. 126). Feedback during training (of any type) is important for three primary reasons: (1) it can help to increase motivation by showing that there is a discrepancy between current performance and the desired level of performance, (2) it can reduce uncertainty (which can take up resources) of how someone is performing, and (3) it can help someone learn how to correct mistakes (Davis, Carson, Ammeter, & Treadway, 2005). For these reasons, feedback is a necessary component for training.

Feedback can be provided by an external source, or it can occur implicitly (Narciss, 2008). Implicit feedback happens "naturally" without any additional outside information. For example, suppose an individual was learning how to search for and identify IEDs in a virtual environment. If the individual misses an IED, it may detonate and kill several members of his or her unit. This would be an example of implicit feedback in that the individual's action (or inaction, in this case) caused an event in the environment to occur. The trainee would then need to make the necessary mental connections between the decision and the outcome. Conversely, explicit feedback occurs when guidance is given outside the context of the task and gives the individual information regarding aspects of his performance (Kluger & DeNisi, 1996). For instance, imagine an individual is going through the same scenario described above, except that the missed IED does not explode. Instead, the instructor points out where the missed IED was so that the trainee can make connections, learn the material, and consequently find the IED next time. Explicit feedback is the type of feedback that is explored in the current research.

At a very basic level, explicit feedback can function either in a confirmatory or a corrective capacity. Both confirmatory and corrective feedback interventions are important because they address different things during training. Corrective feedback focuses on correcting errors, while confirmatory feedback focuses on reinforcing correct answers or actions (Mory, 2004). Kulhavy and Stock (1989) suggest that confirmatory feedback should function to strengthen the response or action so that it is performed consistently over time. For example, in confirmatory feedback, individuals are praised after they perform a desired action or give the appropriate response; this serves to increase and maintain performance and morale (Mory, 2004). Conversely, corrective feedback should function to highlight the error or incorrect action, replace the incorrect action with the appropriate response, and reinforce the correct action so that individuals will be more likely to respond appropriately in the future (Kulhavy & Stock, 1989).

Unfortunately, not many people are able to perform a task correctly on the first try. When people inevitably make errors during training, this presents the instructor with an excellent opportunity to provide guidance which can help individuals recognize mistakes and correct any misunderstandings (Mory, 2004). In this respect, one of the critical functions of feedback is to provide corrective information so that an individual becomes aware of his or her errors, learns from the mistakes, and gains a deeper understanding of the information. Mory (2004) points out that, “Because the correction of errors appears to be where feedback has its most promising effects, researchers should continue to examine ways in which to manipulate feedback to maximize this outcome” (p. 758). One way to manipulate corrective feedback is to change the feedback content, or the information included. Feedback that is intended to be corrective must contain at least one element; it must verify the correctness of an individual's answer or performance. However, often this is combined with additional information for the individual to direct his or her attention to the particular errors (Mory, 2004).

## Manipulating the Content of Feedback

Feedback content refers to the type of information that is included in a feedback message, and it can be addressed in terms of whether the feedback contains verification of information, elaboration of information, or a combination of both. Verification refers to information about the correctness of an answer (e.g., an overall percentage score; Shute, 2008). This serves to evaluate the trainee on his or her performance. Outcome feedback is an example of feedback that only incorporates verification information. On the other hand, elaboration refers to instructive information that helps guide the trainee towards an end goal (Shute, 2008). Feedback that incorporates both verification and elaboration elements is referred to as formative feedback. Outcome feedback and formative feedback are both addressed in more detail.

### *Outcome Feedback*

Outcome feedback includes only verification information and is sometimes referred to as knowledge of results (KR) or knowledge of performance (KP) feedback. It provides information regarding an individual's effectiveness in completing a task (Kluger & DeNisi, 1996; Narciss, 2008). For example, outcome feedback may include a message that says, "your answer/action was correct/incorrect" or "you scored 75% on that exercise." While outcome feedback is commonly used in training systems, Kluger and DeNisi (1996) assert that feedback improves the learning process when it focuses an individual's attention on "…discrepancies between the hypotheses (standards) regarding the details of task performance and the outcomes of acting on these hypotheses" (p. 265). Outcome feedback does not provide all of that information. Kluger and DeNisi (1996) go on to further state that, "If the [outcome feedback] is not accompanied with cues helping to reject erroneous hypotheses, it may cause the recipient to generate a multitude of hypotheses that can reduce consistency and hence decrease performance" (p. 265). In other words, outcome feedback by itself cannot adequately support an individual. On the other hand, formative feedback (i.e., corrective information in addition to outcome feedback) helps an individual learn because he or she is able to monitor his or her own progress and learn how to improve performance on the task.

Kluger and DeNisi's (1996) position is supported by empirical evidence in a recent simulation-based study conducted by Astwood, van Buskirk, Cornejo, and Dalton (2008). These researchers conducted research in which they examined performance effects of four different feedback conditions: process feedback (i.e., step-by-step instructions about how to perform the task), outcome feedback (i.e., knowledge of results, "You were correct XX% of the time"), normative feedback (i.e., information about performance relative to others), or no feedback (Astwood et al., 2008). The participants were trained to perform a simulated task as part of the Fire Support Team (FiST) to disable enemy targets. Results showed that trainees in the process feedback group performed significantly better than those trainees in the outcome, normative, and no feedback groups (Astwood et al., 2008). In addition, Astwood and colleagues (2008) found that performance did not differ significantly between the no feedback group and the outcome feedback group.

In another study, a computer-based training program using a simulated water purification plant was used to examine the effects of several different kinds of feedback (Gonzalez, 2005). There were several different groups of participants: a control group (i.e., outcome feedback, as expressed by the numbers of gallons remaining in the water purification system at the end of the exercise), a feedback group (i.e., detailed performance and outcome feedback was given), a self-exemplar group (i.e., participants saw a replay of the exercise they just completed and were asked to analyze their decisions), a feedback-exemplar group (i.e., received detailed feedback and could also replay the exercise they just completed), and an expert-exemplar group (also called feedforward feedback; replaying how an expert would complete the exercise; Gonzalez, 2005). Participants completed exercises where they made decisions to activate and de-activate different water pumps in order to fill a certain number of water tanks before time ran out. The expert-exemplar group showed the most performance improvement, while the other feedback groups did not show any performance benefits over the control group (Gonzalez, 2005). Here, outcome feedback proved to be an ineffective feedback strategy. These results imply that more elaborate feedback (also known as formative feedback) is more effective than outcome feedback. The benefits of formative feedback are discussed in more detail.

### *Formative Feedback*

Some researchers believe that feedback should be formative, incorporating the elements of both verification and elaboration (Kulhavy & Stock, 1989). While outcome feedback is just intended to alert an individual about his or her performance on the task, “the main aim of formative feedback is to increase student knowledge, skills, and understanding in some content area or general skill (e.g., problem solving), and there are multiple types of feedback that may be employed toward this end (e.g., response-specific, goal directed, immediately delivered)” (Shute, 2008, p. 156-157). Formative feedback gives more specific, informative guidance that can serve to modify the behavior or responses of the individual (Mory, 2004).

Research has demonstrated that formative feedback has positive benefits on training. For instance, Gilman (1969) showed feedback that was more elaborate than outcome feedback contributed to more positive performance on science-related tasks. Gilman (1969) compared several different kinds of corrective feedback used in computer-aided instruction that taught general science concepts to University students. Students were assigned to one of five feedback groups: (1) no feedback, (2) feedback telling students if they were “correct” or “wrong,” (3) feedback showing the correct response, (4) feedback appropriate to the student’s response, and (5) a combination of feedback groups 2 – 4. The groups where students were shown the correct response (3 – 5) performed significantly better than groups 1 or 2. The combination group that received the most information about the task (5) showed the best retention of information. The most detailed formative feedback group (group 5) exhibited the highest levels of learning and performance.

Graesser, Chipman, and King (2008) suggest that “a test score alone is adequate feedback for informing the learner on how well they are doing but is not useful for clarifying specific deficits in knowledge or skill. We need a better understanding of the conditions under which the learner benefits from feedback in the form of correct answers, why correct answers are correct, identification of misconceptions, explanations of the misconceptions, and other forms of elaboration" (p. 214). For this reason, one aspect of the current research examines how formative feedback (which includes both verification and elaboration) may provide training benefits over simple outcome feedback. While formative feedback has been shown to have learning benefits, another issue regarding the content of the feedback arises. Formative feedback elaborates on mistakes that a person makes during training; however, how specific does this elaboration need to be to maximize learning?

## Feedback Specificity

The corrective information included in formative feedback can differ in the level of information that it contains (feedback specificity; Shute, 2008). On one hand formative feedback can be detailed and can tell the individual exactly how to fix the problem or perform the task (Shute, 2008). Very detailed feedback is inherently longer and more specific than general feedback, and it culminates in an explicit answer or solution for the trainee. In other research, this is sometimes referred to as process feedback (Delgado, 2005; Hattie & Timperley, 2007; Astwood et al., 2008), elaborate feedback (Narciss & Huth, 2004; Smits, Boon, Sluijsmans, & van Gog, 2008), directive feedback (Sanders, 2005; Shute, 2008), and feedback that has a high level of specificity (Shute, 2008; Davis et al., 2005; Goodman & Wood, 2009).

On the other hand, formative feedback can be very general and make very conceptual, broad suggestions to gently guide students towards an end goal (Shute, 2008). General feedback often includes hints and minimal information that nudge a person in the right direction without explicitly giving the answer to the problem. In other research, this is similar to global feedback

(Smits et al., 2008; Davis et al., 2005), conceptual feedback (Hays et al., 2009, Cagiltay, 2006), facilitative feedback (Shute, 2008), and hints-based feedback (Shute, 2008). Detailed and general feedback represent the polar opposites in terms of specificity, and levels of feedback specificity can fall anywhere between these two. Feedback specificity may have a major impact on how effective feedback can be during training, so it is a very important aspect of formative feedback that should be considered.

### *Advantages of Adapting Feedback Specificity During Training*

One of the major areas of feedback research is to determine the most appropriate feedback (in terms of specificity) for any given point in time during training. How specific does feedback need to be during the beginning of training as opposed to the end of training? Research has suggested that the appropriateness of feedback content may change as a person learns and that providing certain feedback at the right times may lead to increased performance (Reiser, 2004; Pea, 2004). Therefore, adapting feedback specificity to each individual may promote learning. The idea of adapting instructional components to individual students can be traced back to early research on human tutoring (Anania, 1983; Burke, 1984; Bloom, 1984). Tutoring refers to instruction given by a human tutor to an individual student, where there is constant reinforcement and corrective guidance tailored specifically for an individual (Bloom, 1984). A human tutor is extremely effective because he or she is able to gauge prior knowledge, potential misunderstandings, and confusion and then adapt the guidance and feedback to enhance the learning experience (Anania, 1983).

Two studies in particular have directly compared the benefits of tutoring and adapting to each individual with conventional instruction and mastery learning (Anania, 1983; Burke, 1984). Conventional instruction refers to group instruction where no additional guidance is given to students; they are simply tested at the end of the entire lesson (Anania, 1983). This is considered to be a low quality instructional technique because, "conventional group instruction cannot provide optimal qualities of instruction for all members of the group because of individual differences in students' cognitive and affective entry characteristics" (Anania, 1983, p. 1). In conventional instruction, instructors do not adapt learning cues to each individual, and consequently errors and misunderstandings can easily occur (Burke, 1984). Mastery learning was much like an enhanced conventional classroom. Students were tested periodically to make sure they had achieved a certain criterion level (80%); those who did not meet this criterion after initial instruction were given additional guidance to reinforce learning objectives in areas where performance was lacking (Anania, 1983). Finally, in one-to-one tutoring, each student was paired with a tutor. Anania (1983) compared learning outcomes when students were assigned to a conventional instruction group, a mastery learning group, and a one-to-one tutoring group using different content areas (probability and cartography) as well as different grade levels ($4^{th}$, $5^{th}$, and $8^{th}$ grade students). Results showed that across both the content areas and grade levels, those students who were in the one-to-one tutoring group had the highest levels of achievement, and the students in the mastery conditions performed better than those students in the conventional classroom groups (Anania, 1983). In another similar study, Burke (1984) also directly compared student performance in conventional classrooms, mastery learning classrooms, and tutoring environments. Students were randomly assigned to a learning group. Overall, tutoring led to significantly better achievement and performance (Burke, 1984).

In a seminal article on the benefits of one-to-one tutoring, Bloom (1984) discusses the importance of the research findings of Burke (1984) and Anania (1983). Both studies found that one-to-one tutoring led to significantly better achievement scores than the other two methods of instruction (Anania, 1983; Burke, 1984). In fact, Bloom(1984) points out that the students receiving tutoring in these studies performed two standard deviations (sigma) better than the students in the conventional classroom and one standard deviation better than those students in the mastery learning classroom. Therefore, one-to-one tutoring appears to be the best type of instruction because it addresses the needs of each student (Bloom, 1984). However, how can this personalization be transferred to situations in which groups of students can be enabled to learn as effectively as if they were receiving one-to-one human tutoring? Bloom (1984) refers to this as the “2 sigma problem,” which is the issue of whether instructional designers can figure out a way to incorporate the characteristics of one-to-one tutoring so that groups of students are able to perform as well as those who are actually receiving tutoring. This idea is applicable to the future of SBT because these training systems are designed to train many people (sometimes simultaneously). In addition, most current SBT systems do not adapt to each user. Based on Bloom’s (1984) suggestion, designers of SBT systems need to examine how to incorporate adaptive elements so that these simulations can successfully emulate one-to-one tutoring. The current research focuses on how to administer feedback that adapts to the needs of an individual, much like a human tutor would.

Recent research has attempted to examine feedback specificity and how to sequence different kinds of feedback during training in order to better adapt to the individual. Some research has adjusted feedback specificity using a time-based approach. Here, the feedback specificity changes at pre-determined times during training, usually following some sort of learning model. Unfortunately, this is not truly adaptive feedback because all trainees follow the same feedback sequence, and no one receives a unique feedback experience. Conversely, feedback specificity should be adapted to an individual based on ongoing performance and knowledge level (using a mastery-based approach), which more closely resembles characteristics of one-to-one tutoring. Mastery-based adaptive feedback is dynamic and relies on continuous performance assessments for each individual trainee. The performance measure is used to infer knowledge level and understanding. Feedback is then adapted to an individual's performance to accommodate his or her learning, which is very similar to the interaction between a human tutor and a student. It is possible that this mastery-based approach may be critical in addressing the "2 sigma" issue in SBT, although very little research utilizes this approach.

The Cognitive Load Theory (CLT), along with the theory's application to the interpretation of the Expertise Reversal Effect (ERE) suggest that mastery-based adaptive feedback is an effective approach. However, even though the literature supports the mastery-based approach, many training systems and research designs continue to incorporate a time-based feedback strategy because this is simply easier to design and execute. The current research implements adaptive feedback using a mastery-based approach, distinguishing it from most of the existing literature.

*Time-Based Feedback*

With a time-based approach to administering feedback, researchers can examine how people perform and learn when given different feedback at different times over the course of training. Time-based feedback does not change dynamically based on an individual's performance or state characteristics during training. Instead, the feedback content changes due to pre-set conditions (e.g. time, or task), determined by the instructor. The feedback specificity is switched based on an assumption that at a certain point in time, all trainees should have mastered the task to some degree. For example, imagine that an individual is required to complete four training exercises. The individual receives feedback after each training exercise, which can be either detailed feedback about the steps necessary to achieve a goal or general feedback about the task. Now suppose the instructor chooses to give detailed feedback to individuals on the first and second exercises and general feedback on the third and fourth exercises. Thus, regardless of performance, the feedback will follow this predetermined sequence across exercises for all individuals. Time-based feedback is similar to more traditional instructional methods because the individual learner's unique needs are not taken into account; instead all students are treated the same way, and individual differences are ignored. In feedback research, time-based feedback is similar to the implementation of scaffolding feedback (Jones & Fleischman, 2001; Sharma & Hannafin, 2007), fading feedback (Goodman & Wood, 2009; Jones & Fleischman, 2001; Kester & Kirschner, 2009), reverse fading feedback (Goodman & Wood, 2009), and sequencing feedback (van Duyne et al., 2001; van Gog, Paas, & van Merriënboer, 2008).

Frequently, feedback research will use this time-based approach (Goodman & Wood, 2009; van Duyne et al., 2001). For example, Goodman and Wood (2009) examined how to sequence feedback interventions of varying degrees of specificity. They examined how fading feedback specificity versus increasing feedback specificity affected performance during simulated work team management training. In this study, faded feedback was implemented by presenting more feedback initially and fading the amount of information contained in the feedback over time (Goodman & Wood, 2009). Increased feedback was implemented in the opposite way: more specific feedback was given gradually as training continued. Participants (who were novices) were trained to perform management decision making in a simulated furniture factory over a total of 18 trials. Participants were randomly assigned to either the fading feedback group or the increasing feedback group. For the faded feedback group, very specific feedback was given during the first six trials. Then moderately specific feedback was given during the next six trials. Finally, the least specific feedback was given during the last 6 trials. For the increased feedback group, the order was simply reversed. Results indicated that fading feedback did not show better transfer than the increased feedback condition, contradicting much of the research that documents the benefits of a faded approach (Goodman & Wood, 2009). However, it is important to note that the authors employed a time-based method of administering the feedback such that all participants in the group received the same sequence of feedback. This did not take into account individual learning rates and knowledge levels.

In another study, van Duyne and colleagues (2001) examined how to sequence feedback to generate an optimal learning experience in a simulation-based radar training system. In particular, the effects of presenting process versus outcome feedback were examined. In this study, process feedback involved giving participants step by step instructions on how to perform the task (i.e., it included very detailed and specific information). Conversely, outcome feedback involved giving participants only information about their performance on the task (i.e., it included less specific information). The researchers were also interested in analyzing whether presenting one form of feedback before the other would be beneficial to learning and performance on the simulated task (van Duyne et al., 2001). Participants were randomly assigned to one of the following feedback sequence groups: (1) process feedback for the first half of training followed by process feedback for the second half of training, (2) process feedback for the first half of training followed by outcome feedback for the second half of training, (3) outcome feedback for the first half of training followed by process feedback for the second half of training, (4) outcome feedback for the first half of training followed by outcome feedback for the second half of training, and (5) or a control group where no feedback was given. For the experimental conditions, the feedback was manipulated at the beginning of training and once again in the middle of the training period. Results showed that participants who received feedback, regardless of condition, performed better than those participants in the control group. However, these results should be interpreted with great care because the researchers used a time-based approach to switching feedback content, and it applied to all learners as they progressed through training rather than focusing on each individual learner's needs.

Making a clear distinction between time-based and mastery-based methods of administering feedback is important. The results of these aforementioned studies do not necessarily demonstrate the successfulness of adaptive feedback techniques because of the method in which feedback was administered. Time-based methods do not take the individual into account. However, mastery-based methods provide a uniquely tailored learning experience which responds appropriately to the constantly changing needs of an individual.

*Mastery-Based, Adaptive Feedback*

One of the primary issues in training technology is incorporating instructional environments and conditions that can account for individual differences in goals and learning abilities because these things may give certain individuals learning benefits over others (Park & Lee, 2004; Shute & Zapata-Rivera, 2007). For example, people may learn at different rates and some may become competent and develop proficiency more quickly than others. Mastery-based feedback (also referred to as adaptive feedback throughout the paper) attempts to address this issue by integrating flexible instructional interventions and strategies that can accommodate individual learning needs while ensuring students acquire the desired skills and knowledge (Park & Lee, 2004). Adaptive feedback involves measuring an individual's performance (e.g. test scores, error rates, success rates, domain-specific knowledge tests, etc.) as an indicator of learning, knowledge level, and proficiency, and measuring performance is a common method used to assess knowledge expertise (van Merriënboer & Sweller, 2005; Chi, 2006). Adaptive feedback changes dynamically in response to on-going performance and demonstrated competency; as performance changes (i.e., as knowledge expertise develops), the specificity of feedback given over the course of training changes. This approach to administering feedback gives each person a unique feedback experience that is personalized just for his or her needs. The specificity of the feedback is responsive to how an individual is doing on the task, which is very similar to the interactions between a human tutor and a student.

To date, very little research has been conducted using mastery-based adaptive feedback. In fact, most of the studies relating to adaptive feedback are, in reality, using a time-based feedback strategy to conduct research. It is questionable whether using these time-based feedback sequences actually generates the same learning advantages that adaptive feedback attempts to achieve (Shute & Zapata-Rivera, 2007). Examining mastery-based adaptive feedback more closely is important because it may provide an answer to Bloom's (1984) "2 sigma" problem of how instructional designers can increase student performance during instruction to the level of performance seen in one-to-one tutoring environments. Fortunately, several theories in the literature present basic guidelines for implementing mastery-based adaptive feedback.

## Theory-based Support for Adaptive Feedback

Assessing performance and dynamically tailoring feedback to an individual may prove to be beneficial because it takes into account individual differences in learning. However, what is the best way to implement adaptive feedback in a SBT system? For instance, should individuals be given detailed feedback or general feedback at the beginning of training? What performance levels should cue the transitions in feedback specificity? The Cognitive Load Theory (CLT), which incorporates the concept of Working Memory, and the Expertise Reversal Effect (ERE) provide some loose guidelines for creating adaptive feedback in SBT. The assumptions and implications of these theories are discussed in length.

### *Working Memory*

When designing adaptive training systems, working memory capabilities and limitations should be taken into account because these are integral to the learning process. Working memory (WM) includes all of the processes necessary to temporarily store, manipulate, and integrate new information in short-term memory and then transfer information to long-term memory (LTM), all of which require WM resources (Baddeley, 2000). First, external stimuli are attended to in sensory memory (where they may activate prior knowledge in LTM and hence utilize less cognitive resources) and transferred into WM (Kalyuga, 2009). In WM, the information is integrated and mental representations are constructed and remembered if adequate attention and resources are devoted to the task (Kalyuga, 2009).

Several different models of WM exist, but the Baddeley and Hitch model (1974) delivers the most comprehensive one. It proposes that WM has three structural components with limited processing capacity (Baddeley, 2000). The three components include two slave systems (that are responsible for temporary maintenance of information) and a central executive (that is responsible for supervising information integration and for coordinating the slave systems). One of the slave systems is the phonological loop, and it stores phonological information (e.g. the sounds of language) and prevents its decay by continuously articulating its contents in a rehearsal loop. Conversely, the other slave system is the visuo-spatial sketchpad. This system stores visual and spatial information and is used for constructing and manipulating visual images. The third component is the central executive. The central executive has limited resources available for processing information (e.g. directing attention to relevant incoming information, suppressing irrelevant information and inappropriate actions, coordinating cognitive processes when more than one task must be done at the same time) and storing information temporarily during processing (Kalyuga, 2009). In fact, the central executive trades off resources in order to perform these tasks; if the cognitive resources needed to perform both processing and storage tasks exceed a certain threshold, an individual's performance can suffer (Kalyuga, 2009).

Baddeley (2000) extended the original model by adding a fourth component, the episodic buffer, which holds representations that integrate phonological, visual, and spatial information, and possibly information not covered by the slave systems (e.g., semantic information, musical information). Baddeley's model of WM provides the foundation for the CLT, which is an instructional theory that can be used to guide the design of feedback in SBT systems.

### *Cognitive Load Theory*

According to the CLT, the right type of training can reduce unproductive sources of cognitive load that may hinder learning and increase productive sources of cognitive load so that an individual can learn more efficiently (Clark et al., 2006). Most researchers recommend shaping instruction around Miller's (1956) magical number 7 plus or minus 2 to avoid working memory overload. Clark and colleagues (2006) propose that, "[the] Cognitive load theory is the 21st century update to that maxim" (p. xvi). Just as Miller's (1956) magical number 7 plus or minus 2 has proven to be extremely useful in instructional design, CLT may be key in shaping and integrating adaptive feedback in SBT systems.

CLT applies to all instructional content, all delivery of instruction, and all learners (Clark et al., 2006). CLT is based on the idea that WM has limited resources and that LTM is limitless (Kalyuga, 2009). The theory suggests that the ultimate goals of instruction should be schema (plural: schemata) construction and automation, which is achieved when a schema is transferred to LTM (Kalyuga, Ayres, Chandler, & Sweller, 2003). A schema can be defined as the categorical rules that a person uses to make sense of the world, and schemata are central to the CLT. For example, most individuals have constructed a kitchen schema that tells them what a normal kitchen looks like. It is likely that a kitchen will have a refrigerator, but it is highly unlikely that a kitchen will have a sofa and loveseat. Another example may be an office schema that tells an individual that computers and desks are likely but that a piano does not fit in that environment. These schemata direct people to the important things in a situation and alert them to things that are out of place or abnormal. They can influence what we pay attention to, and they help us better understand the world (Reisberg, 2006). Our existing schemata (like the kitchen schema example used earlier) can be modified and revised continually by integrating new information (Widmayer, 2007). Consequently, when an individual processes new information in WM, a new schema can be constructed, or that new information can be used to modify an existing schema. A schema is treated as a solitary, single unit of information in WM instead of the many smaller pieces of information that make up the schema (Kalyuga, 2009). In this way, schemata require less processing resources in WM.

A schema applies to specific situations and contexts, and the complexity of a schema differs between experts and novices. For example, an expert chess player will have more complex, well-constructed schema about the possible movements available on a turn and consequences of those movements. Therefore, the expert will be able to function exceptionally well in that domain. On the other hand, a beginner may not have a schema for chess; he or she may not realize that certain movements may have negative future consequences. Therefore, a beginner may have trouble playing the game (Widmayer, 2007). Constructing schemata in WM and transferring them to LTM is necessary to move a novice learner to an expert in the knowledge domain. The assumptions of CLT are all based on this notion of schema construction.

CLT assumes several things about human memory, knowledge construction, and how an individual processes novel sensory information (van Merriënboer & Sweller, 2005; Sweller, 1988). The first assumption is that WM is limited in its resources, and long term memory (LTM) is almost limitless. According to Miller (1956), WM capacity is the magic number seven plus or minus two bits of information. This is not very much capacity, particularly if an individual is a novice and the information is brand new. An individual has even less capacity when learning new information and trying to integrate it, especially for complex tasks (Van Gog et al., 2008). Along the same lines, the second assumption of CLT is that the schemata are created in WM and transferred to LTM. These schemata reduce the cognitive resources necessary for processing information. For example, if a novice learner is presented with many bits of new information, WM is needed to both process the new information and also to form connections between them (i.e. construct schemata). Because humans have limited resources available in WM, this may create bottlenecks and not allow novices to successfully construct schemata or complete the task. On the other hand, an expert in the knowledge domain must simply retrieve his or her previously constructed schemata from LTM to perform the task. Therefore the expert is not as limited by WM capacity.

The third assumption of CLT is that WM helps construct these schemata by processing new information, combining information, and rehearsing it so that it can be transferred to LTM (i.e. it can be automated). If there are too few resources in WM, cognitive overload can occur, where new schemata cannot be created or transferred to LTM (Ayres & van Gog, 2009). When schemata are formed, they help alleviate some of the resource issues in WM. The main idea proposed by the CLT is that the effectiveness of instruction can be increased dramatically by taking into account WM and the associated limitations of resources. Instruction should not demand more resources than are available in WM because this hinders the learning process.

According to these three assumptions, schema construction is integral to learning because it reduces the amount of cognitive resources needed to process that information in the future. This reduction in needed resources is extremely important. If a person needs to process a lot of new information, WM resources can become depleted, and the cognitive load that a person experiences can increase. The CLT proposes that there are three types of cognitive load: intrinsic, extraneous, and germane load. Intrinsic cognitive load consists of the difficulty of the task itself, and novice learners have no schema associated with this at first, so information processing will take more resources (Bannert, 2002). This type of cognitive load typically cannot be manipulated by the instructor. The creation of schemata means that intrinsic load is lower and WM has more resources it can devote to new learning information—more learning can take place (Ayres & van Gog, 2009). Extraneous cognitive load is dependent on the actual design of the instruction, which can be changed by the instructor (Bannert, 2002). Lowering extraneous cognitive load is the focus of many instructional design recommendations (van Merriënboer & Sweller, 2005). Germane load involves the process of learning and occurs when WM has enough resources available to process information more deeply and build schema (Bannert, 2002).

In conclusion, the CLT can been used to guide the design of training systems and other forms of instruction. For example, research on the CLT demonstrates that worked examples, or detailed information, during training can reduce cognitive workload (van Merriënboer & Sweller, 2005). In addition, Kalyuga (2006) speculates that presenting the correct forms of guidance and feedback are critical at different stages in the learning process because this can directly affect how well a person can process information in WM and whether or not effective learning will take place.

### *Expertise Reversal Effect*

The Expertise Reversal Effect (ERE) is a phenomenon that is interpreted using the CLT, and it can be applied to technology-based instruction (Sweller, 2008) and to adaptive training systems in particular. The ERE occurs when one form of instruction is extremely beneficial for novice learners while the same instruction has no effect (or even a negative effect) for individuals who have already learned the information (van Merriënboer & Sweller, 2005). According to the ERE, individuals need a lot of information when they are novices and the amount of prior knowledge in the area is low. However, as they become more familiar with the information, presenting individuals with the same specific feedback actually may begin to hinder performance because the information becomes redundant, increases extraneous load for the individuals, and interferes with their learning (Sweller, 2008). Essentially, the methods of instruction and feedback that are most effective for novice learners may become less effective (and perhaps even damaging to performance) as these learners become more knowledgeable (i.e. they become more like experts in that area; Kalyuga, 2007). In other words, "information that is redundant for a more expert learner may be critically necessary for a less expert learner. A novice may need to borrow information from someone else, an expert may not" (Sweller, 2008, p. 377). The proper type of feedback must be given at the right time, depending on the amount of knowledge that has been acquired by an individual. This suggests that feedback should be adapted to the individual as competency increases.

The ERE takes a schema-based approach and is thought to occur due to differences in novices and experts regarding how well-formed their schemata are (Kalyuga, 2009). For example, novices tend to rely on very low-level schemata, which only take into account “surface aspects” of the task (Kalyuga, 2009). Conversely, an expert will trigger higher-level schemata that incorporate more conceptual information (Kalyuga, 2009). Clark and colleagues state that:

> "As a result of their enhanced schemas, experts have significantly different psychological capabilities than novices. Experts are able to tackle complex tasks that overwhelm less experienced workers. When learning new skills in their domain, experts are enabled by their rich storehouse of schemas to process much larger amounts of information as well as to guide much of their own learning processes. Novices, in contrast, lack such schemas and therefore need learning environments that compensate for them. Well-designed learning environments for novices provide schema substitutes by optimizing the limited capacity of working memory in ways that free working memory for learning" (Clark et al., 2006; p. 32).

Kalyuga (2006) points out several differences between the cognitive architecture of novices and experts. First, if a novice does not have external guidance, he or she may use bad strategies or weak, time consuming approaches that can expend more cognitive resources (Kalyuga, 2006). This also means that the creation of schemata will be more difficult and a lot slower. On the other hand, experts have already organized their schema in LTM, and these are automatically referred to when seeing a familiar problem or situation. This makes experts much more efficient at processing information with less cognitive load (Kalyuga, 2006). These differences affect how advantageous different kinds of guidance are when learning new things. “The expertise reversal effect suggests that the detail provided in technology-based instruction should be determined by the knowledge base of the learners. Details that are essential for novices may be redundant for more expert learners. Thus, technology-based instruction must be constructed so its specifications changes with changes in expertise" (Sweller, 2008, p. 377). Therefore, basing the implementation of adaptive feedback on the ERE may help to maximize the benefits of training. A training system that ignores the limitations of working memory in dealing with novel information and also disregards the changing cognitive needs as people become more familiar with information is likely to be completely ineffective (Kirschner, Sweller, & Clark, 2006).

# CHAPTER THREE: LITERATURE REVIEW

Research has suggested that the Cognitive Load Theory (CLT) and the Expertise Reversal Effect (ERE) are important to consider in the design of instructional components. Further, they provide guidance for the implementation of adaptive feedback in SBT systems. Studies supporting and contradicting the CLT and ERE are discussed.

## Cognitive Load Theory Research

Feedback in SBT systems is an area that lacks substantive empirical research. Feedback can consist of information that is very minimal (for instance, outcome feedback gives students only knowledge of results or performance) or very specific and detailed (such as step-by-step processes to achieve the end goal), and everything in between. The existing empirical studies on feedback content often refer to the CLT to explain findings that suggest that more specific and detailed feedback is better for learning. Most research on the content of feedback have revealed that more detailed feedback is especially beneficial for novice learners (Kalyuga et al., 2003; Kalyuga, 2006; Kalyuga, 2007; McLaughlin et al., 2008; Moreno, 2004; Reiser, 2004; Renkl & Atkinson, 2003).

Reiser (2004) suggested that there are several challenges for novice learners that need to be addressed when designing training components (including how specific to make feedback). First, novice learners need to know how to get to the goal when learning unfamiliar information and tasks (Reiser, 2004). This means that they need explicit strategies to help guide them to the goal. Second, novice learners need help connecting information to the task and then generalizing to other tasks (Reiser, 2004). Novices may be thinking at a surface level instead of really reflecting on and understanding the underlying concepts of the task, and they may require help constructing a schema. Third, novice learners may be over-confident in their abilities (i.e. falsely thinking they understand something) while not always performing effectively (Reiser, 2004). Reiser (2004) implies that more detailed guidance is needed to help novices achieve learning objectives, and he believes that general guidance may be detrimental when students are presented with very unfamiliar tasks:

> "Calling students' attention to and requiring use of unfamiliar strategies may work against a system's usefulness for guiding students' investigations. It may require additional reasoning steps that work counter to the structures intended to be useful. Or, if the strategies are unfamiliar enough and students cannot make the connections to their own ways of thinking, they may use the systems' structuring improperly or superficially. For example, despite careful crafting of prompts to guide students' work, students may treat the software environment as "just another worksheet" and ignore the fine distinctions in the systems' attempt to structure the reporting of their work, or may enter minimal answers rather than carefully considering what is needed" (p. 296).

Reiser (2004) emphasizes the difficulty that a novice may face when training to perform a new task. Giving novices very directive support during training may help to offset the disadvantages of being a novice learner, as demonstrated by other research as well. For instance, Gilman (1969) conducted very early research on feedback methods used in computer-based instruction to teach general science concepts. Students were assigned to one of five different feedback groups: (1) no feedback given, (2) feedback telling students whether they were wrong or right, (3) feedback telling students what the correct answer was (knowledge-of-correct-response), (4) feedback telling students why their answer is correct or incorrect (response-contingent feedback), or (5) a combination of correct/wrong, correct response, and response-contingent feedback. Gilman (1969) found that the more detailed feedback in groups 3, 4, and 5 yielded better performance on post-tests than groups 1 and 2, suggesting that giving students detailed knowledge of the correct response or action can significantly improve performance during training.

Another study examined performance effects due to different types of feedback on a computer-based task where education principles were taught (Waldrop, Justen, & Adams, 1986). Students were assigned to three different feedback groups that differed in the amount of information contained in the feedback messages (Waldrop et al., 1986). The first feedback condition presented students with minimal feedback of “correct” or “incorrect.” The second condition included minimal feedback plus extended feedback; in other words, students received minimal feedback for several exercises. If the students still had not improved after these exercises, they were provided with additional explanations for the correct response. The third condition presented extended feedback to students, which included a detailed explanation of the answers. Waldrop and colleagues (1986) found that the most elaborate (i.e. more detailed) feedback increased the understanding of learners and hence increased their performance on the task significantly more than the minimal feedback. These studies emphasize the importance of presenting an individual with a lot of detailed feedback in simple computer-based applications.

In another study, Moreno (2004) examined how different kinds of feedback can influence the effectiveness of discovery learning environments, in which individuals figure out principles and learn concepts on their own. While discovery learning environments may be beneficial in promoting deep understanding by actively involving students in the learning process, there is a caveat; according to CLT, novices may have trouble in these kinds of environments because they are required to learn via free exploration, which may utilize a lot of cognitive resources (Moreno, 2004). CLT suggests that giving people detailed support may be beneficial to learning, and Moreno's (2004) research investigated how applying this to discovery learning may also be beneficial. Two experiments were conducted in which botany was taught using a computer training system. The retention of declarative knowledge was measured as well as the effectiveness of explanatory feedback as opposed to outcome feedback that gave people knowledge about the correctness of their answers (Moreno, 2004). The results of these experiments indicated that participants in the explanatory feedback condition scored higher on a transfer test in both experiments. People in the explanatory feedback group also rated the instruction as easier than the people with the outcome feedback. These results offer further evidence that detailed information can benefit novice learners.

Research on feedback content has also been done on more recent advanced simulation systems, with similar findings. Goodman, Wood, and Hendrickx (2004) conducted a study in which they tested different feedback content (specificity) on learning in a managerial decision-making computer-based simulation. They hypothesized that higher feedback specificity (i.e. more detailed information) would enable students to perform better in practice exercises. Goodman et al. (2004) found that more specific feedback increased performance in practice sessions, even though it led to less exploratory behavior.

Another experiment was conducted by Sanders (2005) on the use of feedback to promote learning in a virtual environment where students were taught how to control unmanned vehicles and conduct reconnaissance, surveillance, and target acquisition based on pre-defined rules (the focus was on learning tactical rules). Sanders (2005) found that feedback which identified student errors and offered a corrective action to be taken to achieve the end goal increased the learning of a new procedural skill. In other words, more detailed feedback was better.

In another study, Astwood and colleagues (2008) found that more specific feedback (which they called process feedback and was defined as outlining the steps needed to achieve the goal), increased performance significantly over outcome feedback (i.e. percentage correct) or normative feedback (i.e. scores relative to everyone else) on a computer program that simulated tasks of a Forward Observer (FO) of the Fire Support Team (FiST). For this task, participants (who were novices in this domain) had to disable as many enemy targets as possible before they got too close to the FO position, based on a series of pre-specified procedures.

Finally, Oden (2008) looked at how feedback affected performance in a simulated team search and rescue task. Oden (2008) found that novices that were able to view a video playback of their performance with coaching (i.e. very specific feedback) could adjust their search schemata better and hence, perform better on subsequent tasks. While the simulation research cited here focused on learning information in differing knowledge domains (e.g. military tactics, business management, education), they all suggest that in the context of simulation-based training, more detailed feedback produces better learning and performance.

Other research has taken a slightly different approach in support of CLT, and one in particular has addressed the efficacy of different feedback specificity based on the cognitive capacity of learners (McLaughlin et al., 2008). McLaughlin et al. (2008) conducted research to examine how learners with different levels of cognitive resource capacity may benefit from high levels of guidance during training and how learners who are training on a complex task may also benefit from additional guidance. In her research, the level of feedback (knowledge of correct response vs. knowledge of correct response as well as additional elaboration) and the complexity of the training task (simple vs. complex) were manipulated for people with either high or low cognitive capacity. A logic-gate task was used because it could be manipulated easily to be either simple or complex. Participants were given a post-test immediately after the experiment and then a week later. For knowledge acquisition, the low capacity participants performed better with elaborated feedback, while the high capacity participants performed similarly with both high and low level support (McLaughlin et al., 2008). For the post-test that was given one week later, there was a main effect for the elaborated feedback over the knowledge of results feedback. Consequently, this suggests that for long-term retention and performance overall, more feedback is beneficial for people with both high and low cognitive capacity.

However, not all results from empirical research in this area completely support the CLT. Some studies did not find more detailed feedback to be absolutely superior to less specific feedback. For instance, Delgado (2005) conducted some research to see what types of feedback led to better task performance on a computer-based board game called Mastermind. Delgado (2005) found that feedback did not consistently improve performance; process (i.e. detailed) feedback had no effect on performance, and outcome feedback (which was less specific) caused worse performance over time. In another study, Hays et al. (2009) also found that more detailed feedback did not increase performance. The researchers trained individuals in a simulated bilateral negotiation task, and compared "bottom out" specific feedback that consisted of step by step, direct feedback to conceptual feedback that consisted of hints. Training performance, transfer task performance, and long-term retention (after several weeks) were measured. Hays et al. (2009) found that specific feedback actually led to worse transfer task performance, and they found no differences on training performance or the long-term retention test. Another interesting finding in the feedback research showed that specific feedback was no more beneficial than presenting no feedback at all (Pridemore & Klein, 1995). Pridemore and Klein (1995) examined differing feedback specificity (i.e. none, elaboration, correct answer) in computer-based instruction where students learned how to operate a microscope. Learning and retention for the specific (i.e. elaboration) feedback was not significantly different than the "no feedback" condition. Both showed higher performance than knowledge-of-correct-response feedback. The researchers hypothesized that corrective feedback gave students additional information (causing them to perform better), while no feedback may have motivated students to seek additional information, causing them to perform well and learn more deeply (Pridemore & Klein, 1995). The correct response feedback may not have been the most effective because it gave just the answer to the learner, which did not require students to make connections between the bits of information and actually learn it. The discrepancies between empirical findings regarding the content of feedback and appropriate levels of specificity (i.e. how detailed or general feedback should be) in computer-based training systems are interesting and beg for further research in this area.

### *Implications for the Design of Feedback Content*

Most research based on the CLT has found that feedback can affect the learning process. Proponents of the CLT typically emphasize the need for more detailed feedback during training (McLaughlin et al., 2008). From this perspective, learning improves when detailed feedback is presented because individuals do not have to integrate information themselves, which lowers the amount of resources necessary to process the new information, reduces bottlenecks, and maximizes the efficiency of WM. In other words this theory predicts that novices will perform better when learning if they are supplied with the important connections. Some researchers and instructional designers would disagree with this statement, which is evident because "…unguided or minimally guided instructional approaches are very popular and intuitively appealing" (Kirschner et al., 2006, p. 75). However, these approaches do not take into account human cognitive architecture and ignore empirical studies that have suggested that minimal guidance is less effective (Kirschner et al., 2006). Direct and specific guidance, which gives students information that fully explains the tasks and procedures, support human cognitive architecture and are therefore more effective instructional approaches. In summary, detailed guidance acts as a substitute for missing knowledge and connections that have not yet been constructed. Novices should be presented with feedback that will help them mentally integrate information and build schemata, which reduces extraneous cognitive load (van Merriënboer & Sweller, 2005).

## Expertise Reversal Effect Research

Since novices and experts differ in the ways that they learn and integrate information, it stands to reason that they respond differently to different types of feedback. From examining the CLT, we have learned that detailed feedback (such as worked examples or step-by-step instructions) helps novice learners organize information for schemata that have yet to be created (Kalyuga, 2006). The ERE takes this one step further and suggests that feedback should adapt to an individual's knowledge level in order to help initially create schemata or to direct attention to previously learned schemata.

Research has demonstrated that novices do not learn as well when they are placed in unguided training environments (ICT, 2009). Novices need to be given some degree of guidance when learning new information, especially those involving complex tasks. The content of the feedback should help the novice develop accurate knowledge structures and build schemata in order to better learn the information and eventually become an expert (Cuevas et al., 2004). The ERE recommends that the way in which instruction and feedback are presented needs to change considerably as learners become more familiar with the tasks (Kalyuga, 2006). Hence, it is extremely important to present the right feedback at the right time and to take away elements at the right time (as the individual's knowledge level increases) that would otherwise amplify WM demands.

Specifically, the ERE predicts that learning should greatly benefit from adaptive bottom-up feedback, which is a term used in the current research that refers to a method for implementing adaptive feedback. Using the bottom-up approach, novice learners are initially given detailed feedback. As their competency is demonstrated during training, the learners are then presented with more general feedback. According to the CLT and the ERE, individuals using adaptive bottom-up feedback should perform better and experience less cognitive workload because they receive more detailed feedback initially and less feedback as their performance improves (illustrated in Figure 1). There is a suggested interaction between knowledge level and feedback specificity.

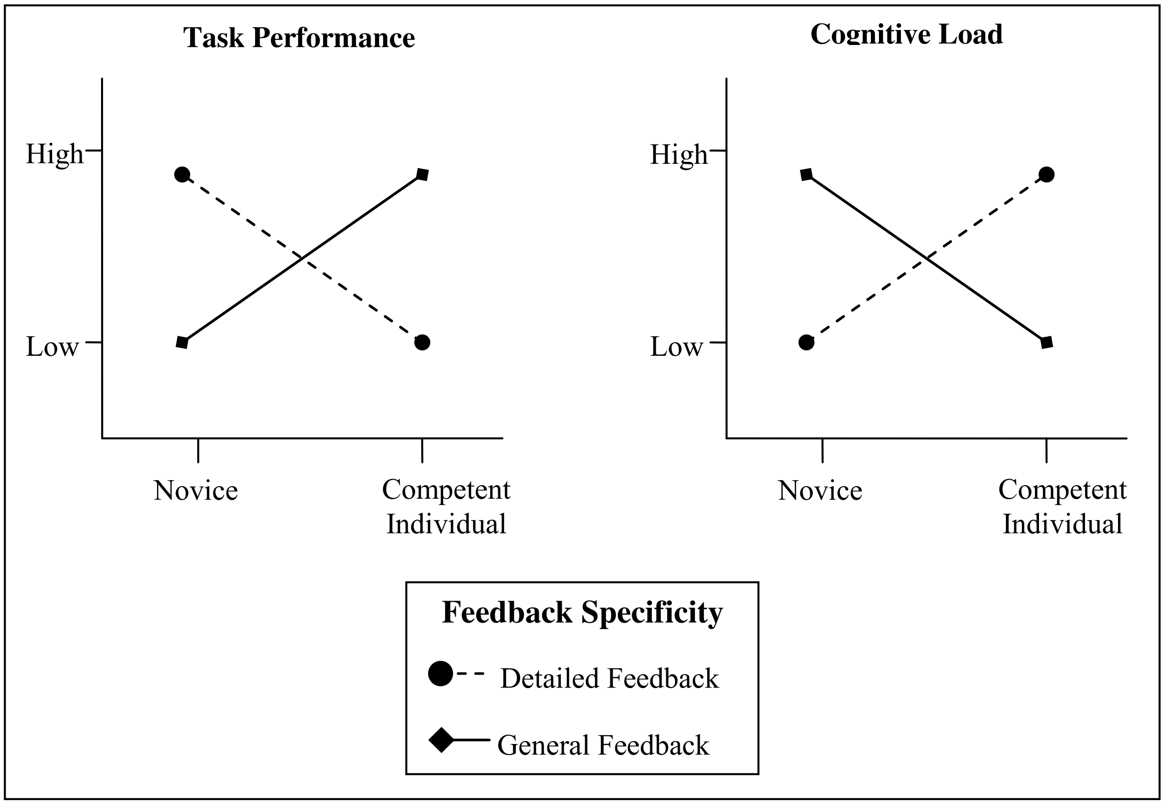



Figure 1. Interaction of Feedback Specificity (Detailed vs. General), Task Performance, and Cognitive Load, according to the Expertise Reversal Effect, adapted from Kalyuga (2009).

Much of the literature corroborates the ERE and suggests that bottom-up feedback is advantageous for novices. For example, in a study by Moreno (2004), computer-based training for botany topics was given to novice learners. Moreno (2004) found that explanatory, more specific feedback led to higher training transfer test scores and lower cognitive load. The learners who received explanatory feedback, as opposed to simple corrective feedback, also rated the computer training more favorably. Another study by McLaughlin and colleagues (2008) found that more feedback helped when the tasks were cognitively demanding and also when the learners were novices and very resource-limited. In addition, van Gog et al. (2008) conducted a study in which participants were given example problems before completing a transfer problem in circuitry and troubleshooting. These example problems contained either process-oriented information (steps to get to the solution) or product-oriented information (presenting the solution without additional information). The researchers had participants go through two practice problems, each followed by a transfer problem. Van Gog et al. (2008) found that presenting process-oriented examples first helped participants construct their schemata and provided more effective transfer on the first transfer problem. In addition, they found that presenting product-oriented examples second led to better performance on the second transfer problem. The authors suggest that an optimal sequence should be providing process-oriented information (i.e. detailed information), then product-oriented information (i.e. general information; van Gog et al., 2008).

While many studies support the effect, the ERE is not always consistently demonstrated. Goodman and Wood (2009) examined whether fading or increasing feedback would affect training performance on a computer-based training simulation for teaching management decisions. In the fading feedback condition, a lot of feedback was given in the beginning of training. After a time, the feedback decreased in the amount of information given. The increasing feedback condition was implemented in the opposite way. Less feedback was given initially, and then after a pre-set time, more feedback was given. Goodman and Wood (2009) found that fading feedback was no better than increasing feedback. While some studies have found no evidence for the ERE, many others suggest that this is an important effect to take into consideration when designing instruction (Kalyuga, 2003).

*Implications for the Design of Adaptive Feedback*

In conclusion, Kalyuga (2006) asserts that novices are at a disadvantage when learning because their WM is extremely limited due to the need to process unfamiliar information. On the other hand, expert learners draw on previously learned schemata in LTM to alleviate some of the cognitive processing taking place in WM. Schemata allow individuals to chunk information (e.g. procedures) and reduce the cognitive demand of tasks (Kalyuga, 2006). Even though using schemata still requires some degree of WM, they can eventually become automatic (through practice), which lowers the amount of WM resources needed to process the schemata. The innate learning differences between novices and experts mean that it is extremely important to design effective feedback to optimize learning at any knowledge level. In effect, the benefits of one type of feedback specificity (e.g., general) cannot simply be generalized to differing levels of expertise. What works for a novice does not always work for an expert, and vice versa.

According to the CLT, if novices are given detailed feedback first (in which they are led step by step through a process), this helps them create schemata for information chunks. These schemata lower the amount of cognitive resources needed to complete the task so that the learner does not get overloaded with the new, unfamiliar information. Essentially, the important connections are made for them. Then, if feedback is subsequently changed to general feedback as they increase in knowledge expertise, this should lead to deeper levels of processing and deeper understanding of the concepts. In this instance, since the learners have schemata already developed, they should not experience cognitive overload. In this way, they are forced to draw on what they have previously learned and connections they have already made to complete the task. The connections are reinforced in this way, and this deeper processing should lead to better retention and performance overall. A training system is likely to be ineffective if the integrated instructional components ignore the limitations of working memory when people process new information or if they overlook the disappearance of those same limitations when people process more familiar information (Kirschner et al., 2006). Even though theory in the literature supports this adaptive bottom-up feedback strategy (where more guidance is presented initially and less is presented as knowledge increases), feedback is often not implemented in this way.

## Current Study

### *Research Question*

The current study investigates two things. First, the research attempts to empirically demonstrate that formative feedback (i.e., feedback that includes both outcome and corrective information) leads to significantly better performance and retention than outcome feedback alone (i.e. feedback that only provides a performance score). This will serve to corroborate prior research studies suggesting outcome feedback is ineffective if delivered on its own. In addition, it will provide support for the CLT, which suggests that detailed corrective feedback is more effective than presenting minimal information. Second, this research attempts to demonstrate that adaptive bottom-up feedback is more effective than non-adaptive feedback strategies (i.e., where feedback content remains either detailed or general throughout the entirety of training) as well as adaptive top-down feedback, which is a commonly used hinting strategy in ITSs (Guo, Beck, & Heffernan, 2008; Schulze et al., 2000).

Adaptive top-down feedback, as given in the present experiment, is conceptually the opposite of adaptive bottom-up feedback. Trainees begin with general feedback (i.e. higher level, conceptual feedback) and transition to detailed feedback if their performance fails to improve during training. From a practical perspective, the inclusion of top-down feedback allows better interpretation of research findings in this study. Suppose the results of this research indicate that adaptive bottom-up feedback is significantly better than non-adaptive feedback groups. These findings would not necessarily provide conclusive evidence that the learning benefit is due to the adaptive nature of the feedback. Instead, the results may be due to the presence of a change in feedback specificity over time. In other words, participants who receive changing feedback over the course of training may have a benefit simply because feedback changes and not because it was adapted to each individual. Perhaps the changes in feedback specificity function to re-direct and re-focus the attention of trainees, thereby enhancing performance. Conversely, people in the non-adaptive feedback conditions only receive the same type of feedback throughout the training process. Perhaps the constant nature of the feedback does not hold their attention. Adaptive top-down feedback was included in the experimental design to address this issue and aid in the interpretation of findings. By adding the adaptive top-down condition, any advantage of the adaptive bottom-up feedback can be attributed to the particular method of implementing adaptive feedback (i.e., bottom-up vs. top-down), not the presence of a change in feedback content.

Not only was adaptive top-down feedback included to help interpret the research findings, but it was also included because this kind of hinting intervention has been traditionally utilized in ITSs (Guo, Beck, & Heffernan, 2008). The theoretical underpinnings for adaptive top-down feedback come from the Vygotsky's notion of the zone of proximal development (ZPD), which reflects the distance between what an individual can accomplish on his or her own and what he or she can accomplish with the assistance of a more knowledgeable person (Kjellin, 2005; Puntambekar & Hubscher, 2005). From this perspective, instruction and feedback should be designed to create learning experiences to support the gradual development of skills and knowledge to the point where students can perform on their own without help (Borthick, Jones, & Waikai, 2003).

One instructional strategy grounded in the assumptions of the ZPD is scaffolding, which is instructional support that helps students perform at higher levels than they can perform on their own without assistance (Bull et al., 1999; Kjellin, 2005). "What the [instructor] does is to probe the student and find out what is not known and then through hints or provision of structures, e.g., advance organizers, shows the learner how the new information can be related to the old" (Bull et al., 1999, p. 241). After initial training or instruction, the goal of scaffolding is to assist as little as possible and only intervene when there is major difficulty or the task cannot be completed. In other words, only when an individual fails to demonstrate a skill or perform at a certain level should the instructor step in and assist the student (Bull et al., 1999).

Many kinds of scaffolding exist, including: (1) giving explanations when individuals do not understand information; (2) using extensive worked examples; (3) using think alouds describing an individual's thinking process; and (4) providing hints or prompting to nudge the individual in the right direction (Bull et al., 1999). Top-down feedback incorporates the idea of scaffolding through the use of hints and progressively more detailed feedback if an impasse is reached. As Murray and Arroyo (2002) state, "we want to give assistance in order to keep the learner at their leading edge—challenging but not overwhelming them" (p. 2). The hints should not be too difficult or too easy, and they should keep an individual's interest while avoiding boredom or excessive confusion (Murray & Arroyo, 2002). If a person is confused or the problem is too difficult, then more directive feedback is needed. Scaffolding is important in the sense that it supports an individual (i.e., helps him or her accomplish a task) and also continues to actively engage that individual in the learning process and challenge the learner (i.e., an individual learns from the effort and experience he or she puts into the task; Reiser, 2004). The student should be provided with just enough support to accomplish the goal and help him or her reflect on the information (Puntambekar & Hubscher, 2005). Reflection is important, and research has indicated that students reflect more on the information when they receive generic prompts than when they receive more specific and directive prompts (Puntambekar & Hubscher, 2005).

An example of an instructional system that utilizes a top-down feedback approach is *Animalwatch*, which is an ITS that teaches basic arithmetic through word problems about endangered animals (Murray & Arroyo, 2002). In this program, if a student enters a wrong answer, the system provides assistance through progressive hints and suggestions. *ELECT BiLat* (Enhanced Learning Environments with Creative Technologies for Bilateral negotiations) is an ITS that teaches culturally appropriate negotiation skills and administers feedback in much the same way as *Animalwatch*—through progressively more detailed hints (Hays et al., 2009). Essentially, the top-down feedback used in the current research functions in a similar way. It initially provides hints (general feedback) to challenge students and force them to exert more effort; in this way students strive to learn and perform with minimal assistance. However, if an individual cannot complete a task on his or her own or experiences great difficulty in completing it, more detailed and explicit feedback is provided to help that individual overcome the impasse. Including this top-down feedback strategy allowed direct performance comparisons to be made with the theoretically-based bottom-up feedback strategy, which is not as widely implemented.

In summary, a total of four formative feedback groups were included in the current research: a constant detailed feedback condition (referred to throughout the rest of the study as "detailed"), a constant general feedback condition (referred to throughout the rest of the study as "general"), an adaptive bottom-up condition (referred to throughout the rest of the study as "bottom-up"), and an adaptive top-down condition (referred to throughout the rest of the study as "top-down"). Each formative feedback group contained a verification component (outcome feedback, which was the performance score) and an elaboration component (corrective feedback that was dependent on the respective feedback group). There was also one control group that received outcome feedback (i.e., a performance score) after each training mission, but no additional corrective feedback.

### *Potential Implications and Relevancy of Research*

SBT systems are primarily concerned with teaching a new concept or task to an individual so that he or she can transfer this knowledge to real tasks or situations. We want trainees to form deep connections and accurate schemata so that when the time comes for them to perform a task without a tutor (and be able to self-monitor their performance), they are able to perform at an acceptable level. Van Merriënboer and Sweller (2005) suggest the need for research on a more dynamic method for providing training, which monitors learning and presents real-time, adaptive instruction. Designing the right components for such a system is a challenge.

In particular, implementing the right kind of adaptive feedback throughout training exercises is a human factors issue that many researchers face in the design of ITSs, SBT systems, and other attempts at adaptive systems. Feedback has been shown to have a positive effect on an individual's performance. However, there is no universal prescription for the right way to implement adaptive feedback.

Adaptive instruction has been useful for human instructors in one-to-one tutoring settings, and adaptive feedback strategies may also be beneficial in computer-based environments with technology-delivered instruction. However, while adaptive techniques have been successful in traditional classroom environments, this may not be true for all computer-based and SBT systems. Too many times, designers find techniques in the literature and attempt to apply them to domains and situations that are substantially different than what has been empirically tested. Simply applying techniques found in the literature does not guarantee that they will be relevant in all circumstances and all systems.

For the development of future adaptive SBT systems, it is a paramount concern to create systems that will respond appropriately to the individual's needs at any given time. The feedback research suggests that both detailed and general feedback have their own merits and can be used successfully to train individuals (i.e. detailed produces good performance initially, general leads to good retention of information). In addition, adaptive feedback may be a useful training component, but most of the research studies in this area implement time-based feedback rather than mastery-based adaptive feedback in their experiments. Empirically tested rules and guidelines for the implementation of adaptive feedback is an urgent need in the field of SBT and adaptive technology development. The current investigation focuses on the efficacy of taking a bottom-up feedback approach from the literature and applying it in a SBT environment.

Finally, a very important concern in the field of training is cost benefits. The cost of developing adaptive technology is usually high, and empirical evaluations of a lot of adaptive training technology are lacking (Shute & Zapata-Rivera, 2007). This means that developers are taking a huge risk with adaptive technology, since there is little empirical substantiation of their effectiveness and the cost-benefit ratio may be too high. Shute and Zapata-Rivera (2007) suggest that controlled studies are desperately needed to gauge whether adaptive technology is truly a cost-effective way to enhance learning. There must be enough value found in adaptive systems to justify the cost. The overall goal of this research is to examine the efficacy of a theoretically-based implementation of adaptive bottom-up feedback. In addition, adaptive feedback may offer quicker training times that could lower overall training costs over time, but again, this needs to be investigated. The findings of relevant research can then be used to create guidelines for adaptive training systems that are timely, cost effective, and efficient instructional tools. This is especially important in military domains because the military is a huge advocate of effective SBT systems (Salas & Cannon-Bowers, 2001).

# CHAPTER FOUR: METHODOLOGY

## Experimental Tasks

The experimental tasks for this research were designed for one individual to perform at a time and included four simulated search and rescue missions. Experimental tasks allowed participants to demonstrate various levels of knowledge. For instance, at the most basic level participants showed that they could operate a computer. At a more conceptual level, participants showed how well they could integrate information and apply rules at different times and in different situations. Participants were required to demonstrate knowledge in three specific learning objectives as they progressed through the missions. They were given feedback following each training mission, based on the feedback group to which they had been assigned.

### *Learning Objectives*

Procedures for completing the simulated search and rescue missions in this study needed to be learned and applied in order to successfully complete the tasks in this experiment. There were three learning objectives (LOs), and they were loosely adapted from Oden's (2008) research using the same virtual environment (GDIS) to train teams. The LOs included: (1) procedures for entering and exiting buildings, (2) procedures for clearing buildings, and (3) procedures for communicating with Headquarters. Under each learning objective, there were several specific procedures that a participant needed to learn. Participants were required to learn declarative information as well as how to recognize cues and how to apply each procedure in different situations. For instance, due to the size and layout of buildings, not all procedures were used every time a building was searched. In addition, the procedures that were used in one situation may differ slightly from the procedures that were used when searching a different building. See Appendix B for detailed descriptions of all search and rescue learning objectives and procedures.

*Missions*

Participants completed four missions during this study. Three training missions were completed in a simulated town (including buildings, trees, and other terrain features). Missions were always presented in the same order because pilot research revealed no significant performance differences due to mission sequence. For each of the three training missions (Mission 1, Mission 2, and Mission 3), the participant was required to search a different set of buildings while following the procedures outlined in the learning objectives. Participants communicated with Headquarters (HQ) using a computerized texting system (the role of HQ was played by a confederate). The task in the three training missions was always the same: search for target items left behind by a missing Alzheimer's patient in a designated set of buildings within the virtual environment with a maximum search time of 10 minutes (See Appendix C for an example of mission instructions). Feedback was given to participants after they performed each of these training missions.

The fourth mission was a transfer mission and occurred in the same simulated town but appeared visually different due to imposed environmental characteristics. Mission 4 involved searching through several buildings for a missing doctor and related target items. Although participants applied the same learning objectives and rules learned to this scenario, most of the target items and visual characteristics of the scenario were different than previous training missions. More specifically, the mission occurred at dusk, in heavy fog, and the target items were relatively unfamiliar to participants. Feedback was not given after the transfer mission. This transfer mission was used because an integral component of search and rescue missions is being able to recognize cues in different environments and apply the appropriate search procedures. Mission 4 was included to ascertain how well participants could transfer search and rescue procedures learned in the prior missions to a visually different environment. This task was designed to be novel for all participants, and while it is not a completely realistic task, it does draw on some of the military operating procedures for search and rescue missions.

### *Feedback Conditions*

Five feedback conditions were included in this study: one functioned as a control group in which only outcome feedback (i.e. a performance score) was given, and four were formative feedback conditions (i.e., corrective feedback was given in addition to outcome feedback). Providing outcome feedback in the form of a performance score was important because research suggests that this can increase motivation, decrease frustration, and ensure active involvement in the task (Jackson, 2007). In addition to a performance score, corrective feedback was provided for the four formative conditions to elaborate on mistakes that were made for each learning objective (LO). Two levels of feedback specificity were used for corrective feedback; messages could either be detailed or general.

In all formative feedback conditions, confirmatory feedback (i.e., feedback that informed an individual that he or she successfully mastered a learning objective) was given for a perfect score on a learning objective. This positive feedback was consistent across all formative conditions. Confirmatory feedback was included to make sure that correct actions were noted, as well as to maintain performance levels and morale (Mory, 2004).

The four formative feedback conditions included: detailed, general, bottom-up, and top-down. In the detailed condition, very specific feedback was given after each training mission for any mistakes made, regardless of an individual's proficiency level. For the general condition, more conceptual feedback was given after each training mission for any errors made, regardless of proficiency level. These two conditions were non-adaptive and did not take into account an individual's knowledge level and mastery of the material. Conversely, the other two formative feedback conditions provided adaptive feedback based on an individual's mastery of the task. Performance was monitored, and feedback for each learning objective was given based on on-going performance assessments. For the bottom-up condition, detailed feedback was given initially; then as competency was demonstrated, general feedback was given. For the top-down condition, general feedback was given initially; then detailed feedback was given if participants failed to consistently improve across training exercises. Both adaptive feedback conditions required criteria for triggering the transition between feedback specificities. Determining exactly what the performance criteria should be proved to be a very challenging task due to the lack of research in this area.

### *Generating Performance Criteria for Adaptive Feedback Conditions*

Performance criteria needed to be set for each adaptive feedback condition in order to cue the changes in feedback specificity (i.e., when to adapt feedback to the individual).

Unfortunately no research was found that documented an established method for generating criteria and implementing adaptive feedback in SBT exercises, where an individual is actively involved in all aspects of the training. However, when consulting the literature for how to determine criteria, related work was found in the area of adaptive automation, where tasks (usually vigilance tasks) are dynamically allocated between a human operator and an automated system (Byrne & Parasuraman, 1996). This research was explored further to determine if the approaches used to set criteria for activating adaptive automation could also be applied to the current study.

The literature on adaptive automation focuses on three different methods for generating criteria (Byrne & Parasuraman, 1996). First, the level of automation can change in response to the occurrence of critical environmental events in an attempt to alleviate workload and improve performance (Scerbo et al., 2001). Therefore, if the events do not occur, changes in the level of automation do not occur. Second, adaptive automation can be activated based on physiological assessments of an individual (i.e., assessments of mental states, emotional states, and cognitive functioning; Scerbo et al., 2001, Byrne & Parasuraman, 1996). For instance, when an operator is experiencing high cognitive workload (overload) or is bored with the task (under-load), the level of automation may change to elicit higher levels of performance (Byrne & Parasuraman, 1996). Finally, criteria can be based on operator performance measurements and tracking performance in real-time, where deviations from a pre-specified performance level trigger changes in the level of automation (Scerbo et al., 2001). This last approach was most appropriate for the current study because feedback was to be adapted based on performance measurements, so research implementing this approach was examined to determine its practicality.

Parasuraman, Mouloua, and Molloy (1996) conducted a study that utilized the performance-based approach to adapting automation. In their study, participants were tasked to monitor an automated system while simultaneously completing other manual flight simulation tasks. Half of the participants were assigned to an adaptive model-based group where the automated task temporarily came under full human control at a designated time, regardless of participant performance on the vigilance task (Parasuraman et al., 1996). In contrast, the other half of the participants were assigned to an adaptive performance-based group where the automated vigilance task only became fully manual for a brief period of time if participant performance dropped below a certain criterion, which in this case was a detection rate of 55%. Results from this study showed that both adaptive groups improved their performance on the vigilance tasks. In effect, Parasuraman and colleagues (1996) demonstrated the viability of using performance-based criteria to trigger adaptation.

In examining literature on adaptive automation, the idea of generating criteria based on performance measures was applicable for the bottom-up condition in the current study. For the bottom-up group, detailed feedback was given to participants until their performance triggered the change to general feedback (i.e., as they demonstrated competency in the task). In other words, all participants in the bottom-up group received detailed feedback after the first mission. However, for the subsequent missions, participants were required to perform above continually rising criterion level in order to receive general feedback. If they did not meet this performance standard, then they continued to receive detailed feedback. The actual performance criteria were generated based on data from a pilot study. Hence, the criteria were not arbitrary numbers and also reflected learning over time, which is an important aspect of training with SBT systems. In the pilot study, 13 participants received detailed feedback across missions for any errors made, regardless of performance level. Thresholds (median scores) were identified that could be used as performance criteria for each of the missions (See Table 1 for specific criteria and a more detailed explanation of the implementation of bottom-up feedback). Using median scores for triggering changes in feedback specificity represented a reasonable level of competency and ensured that at least half of the participants in this condition would receive the feedback manipulation for each mission.

Conversely in the top-down feedback condition, participants were given general feedback unless they failed to improve in their performance over missions, which triggered a change in feedback specificity. This top-down implementation did not require generating specific performance criteria. Instead, an individual's performance was compared to his or her own performance on the prior training exercise to determine if changes in feedback specificity were warranted. If an individual did not demonstrate improvement from one exercise to the next, this triggered detailed feedback. On the other hand, if an individual showed improvement over exercises, he or she was given general feedback.

Generating criteria proved to be difficult because few guidelines outside of the adaptive automation literature exist. This may be one of the reasons why studies on mastery-based adaptive feedback are not readily available in the literature. Even though these criteria suited the purpose of this research, it may not have been the optimal criteria to use. For this reason, future research needs to explore different approaches that can be used to determine criteria for triggering changes in feedback specificity in SBT systems.

### *Implementation of Feedback Conditions*

For the control condition, only a performance score was given. For all the formative feedback conditions (i.e., detailed, general, bottom-up, and top-down), corrective feedback was tailored for each of the three learning objectives (LOs) involved in each mission. The implementations are discussed in more detail.

#### *Detailed Condition: Constant Detailed Feedback*

In the detailed condition, specific feedback was given in addition to a performance score after each training mission (See Table 1 for an outline of all feedback conditions). If a participant made mistakes during a training mission, detailed feedback provided step-by-step information outlining the correct procedures and provided very specific information about the processes required for each of the LOs. It explicitly described what needed to be done to rectify the mistake. For example, if a participant did not walk around the perimeter of a building to search for existing tags before entering, he or she would receive the following detailed feedback after the mission: "Before entering or tagging a building, you should walk around the entire building to make sure it is not already tagged." This detailed feedback was specific for the procedures under each learning objective. Therefore, if participants failed to perform four procedures under a learning objective, they would receive detailed feedback for each of those four procedures. If a participant followed all the procedures for a specific learning objective, he or she received confirmatory feedback (which was identical for all conditions) regarding their performance for that learning objective (e.g., "Great job applying all procedures for entering and exiting buildings!"). See Appendix D for a complete list of the feedback messages used in this research.

Table 1. Feedback Conditions (All Conditions Included an Overall Performance Score).

| Feedback Condition | Feedback Implementation |
|---|---|
| Control group | Outcome feedback (i.e., performance score) was given after each mission. |
| Constant Detailed | Detailed feedback given for mistakes on learning objectives (LOs).<br><br>Confirmatory feedback was given if no errors were made for a particular LO. |
| Constant General | General feedback was given for mistakes on LOs.<br><br>Confirmatory feedback was given if no errors were made for a particular LO. |
| Adaptive Bottom-Up | Detailed feedback was given after Mission 1.<br><br>For Mission 2 and Mission 3, feedback adapted to the participant according to the following criteria:<br><br>If participants scored ≥ the median performance scores of pilot participants shown below, then they received general feedback for that particular LO.<br><br>If the score was < the median scores shown below, then participants received detailed feedback for that LO.<br><br>Confirmatory feedback was given if no errors were made for a particular LO.<br><br>**Mission 2 criterion scores**<br><br>• LO 1: 75<br>• LO 2: 70.83<br>• LO 3: 75<br><br>**Mission 3 criterion scores**<br><br>• LO 1: 81.25<br>• LO 2: 75<br>• LO 3: 80 |

| Feedback Condition | Feedback Implementation |
|---|---|
| Adaptive Top-Down | General feedback was given after Mission 1.<br><br>For Mission 2 and Mission 3, feedback adapted to each participant based on an individual's performance on prior missions.<br><br>For Mission 2, the performance score on each LO was compared to the performance score for each LO in the previous mission (Mission 1).<br><br>For Mission 3, the performance score on each LO was compared to performance score for each LO in the previous mission (Mission 2).<br><br>• If an individual's score on a LO was ≤ the individual's score on that same LO from the previous mission, the participant received detailed feedback.<br><br>• If an individual's score on a LO was > the individual's score on that same LO from the previous mission, general feedback was given.<br><br>Confirmatory feedback was given if no errors were made for a particular LO. |

*General Condition: Constant General Feedback*

For the general condition, general feedback was given in addition to a performance score (See Table 1). If a participant made a mistake during a training mission, general feedback provided a hint to help the person pinpoint where he or she made mistakes, self-diagnose the particular ones, and apply it to the next mission to improve performance. The general feedback was vague and referred back to a learning objective, which cued an individual to think about things learned during training. It never explicitly told participants how to correct problems. For example, if a participant failed to walk around the exterior of a building before entering the building for a search, he or she would get the following feedback after a mission: "Remember to apply the procedures for entering and exiting buildings." This general feedback was given if any of the procedures for that particular learning objective were missed. The general feedback was not tailored for each specific learning objective procedure, as the detailed feedback was. In addition, if a participant followed all procedures for a particular learning objective, they received confirmatory feedback regarding their performance (which was the same across feedback conditions). See Appendix D for a complete list of the feedback messages used in this research.

*Bottom-up Condition: Adaptive Bottom-up Feedback*

For the bottom-up feedback condition, a performance score was given after all training missions. In addition, detailed feedback was given after Mission 1 was completed. Feedback messages for the remaining training missions (Mission 2 and Mission 3) adapted to each participant and were generated based on the performance criteria derived from the pilot study. For Mission 2, the performance scores on each individual LO were compared to the median LO scores of the pilot participants on that same mission (See Table 1). If participants scored greater than or equal to the median scores for each particular LO, then they received general feedback. If participants scored lower than the median scores for each specific LO, then they received detailed feedback. In addition, if no errors were made, participants received a confirmatory feedback message. Therefore, it was possible for participants to receive detailed, general, and confirmatory feedback messages after Mission 2. For example, a participant may receive detailed feedback for learning objective one, general feedback for learning objective two, and confirmatory feedback for learning objective three. The feedback messages for Mission 3 were generated in the same way, comparing performance scores on each LO to the median scores of the pilot participants on the same mission.

*Top-down Condition: Adaptive Top-down Feedback*

In the top-down condition, a performance score was given after all training missions. In addition, participants were given general feedback after Mission 1. Feedback messages for the subsequent training missions adapted to each individual and were generated in the following way. If the current learning objective score was higher than the same learning objective score on the previous mission, then general feedback was given for any errors that may have been committed. In other words, as an individual participant demonstrated competency (by scoring higher than his or her previous score), general feedback was given. However, if the current learning objective score was lower or equal to the same learning objective score on the previous mission, then detailed feedback would be given for any mistakes made. In this way, if participants did not continue to improve their performance scores, they received additional support in the form of detailed feedback. Finally, if no errors were made on a LO, then confirmatory feedback was given. For this condition, it was possible for participants to receive confirmatory, detailed, and general feedback after missions.

*Control Condition*

Participants in the control group only received outcome feedback (i.e., a performance score) after each training mission. They did not receive any additional feedback or support as they progressed through the training missions. This way, any differences in performance between the control group and formative feedback conditions could be attributed to the feedback content above that of simple outcome feedback.

*Timing of Feedback Messages*

The timing of feedback was an issue that was integral to the design and implementation of the feedback conditions. Research on feedback timing has yielded inconsistent findings in the literature. Immediate feedback refers to guidance given immediately when an individual makes a mistake. Delayed feedback can refer to feedback that is given immediately following task completion, or it can refer to feedback given after an entire training session has been completed (similar to an After Action Review). Some studies have found no significant differences between immediate and delayed feedback (Bolton, 2006; Smits et al., 2008). However, other studies show that delayed feedback (i.e. feedback presented immediately after the task is completed; between practice scenarios) is an effective instructional intervention, especially if the task is complex or if it occurs in real-time (Astwood et al., 2008; Corbett & Anderson, 2001; Hattie & Timperley, 2007). Munro et al. (1985) reasons that if you interfere with a complex task, there will be more demand on the learner because attention will have to be shifted to the feedback. If learners have to devote a lot of cognitive resources to the task, intrusions will cause performance to suffer (Schooler & Anderson, 1990). Munro et al. (1985) refer to the attentional demand hypothesis which posits that feedback will disrupt learning if the intervention occurs during task performance, as opposed to feedback that occurs after task performance. Delayed feedback may actually encourage people to monitor themselves and correct their own errors, while immediate feedback only competes for cognitive resources (Schooler & Anderson, 1990). In other words, working memory demands may be higher if an individual has to split attention from task to feedback, and for this reason it may be better to give an individual feedback while they are not performing the task (e.g. reading over it right before they practice the task again). Looking over process information before the task may help to generate schema, which can be retrieved while performing task and be less resource demanding (van Merriënboer & Sweller, 2005). Finally, transfer appropriate processing suggests that simulation training is best when it closely mirrors the real world task it is training for (Bolton, 2006). Because of this, slightly delayed feedback may be better because it does not interfere with the task being learned so that the training task is that much closer to the real task. Hence, the student can learn and remember the information more effectively (Mory, 2004). For these reasons, the feedback in this study was administered immediately following completion of each training mission.

## Experimental Design

### *Mixed Between-Within Design*

This experimental design was a 4 x 5 mixed between-within design with two independent variables. The first variable, mission, was a within-subject variable with four levels (Mission 1, Mission 2, Mission 3, and Mission 4). The second variable, feedback condition, was the between-subjects variable and had five levels (detailed group, general group, bottom-up group, top-down group, and control group). The primary dependent measures included overall performance scores for each of the four missions. In addition, measures of cognitive load, knowledge retention, and subjective opinions regarding the feedback were obtained.

### *Experiment Covariates (CVs)*

Due to the nature of the simulated task, spatial orientation and video game experience were predicted to impact performance on the missions. Consequently, these individual differences were measured and used as covariates (CVs) in the current experiment.

#### *Spatial Orientation*

Research has shown that individuals frequently have trouble with navigation and orientation in virtual or simulated environments (Diaz & Sims, 2003). Many of these navigation problems can be attributed to individual differences, which can affect the performance capabilities and usability of the system (Diaz & Sims, 2003). For instance, spatial ability is an individual difference that is often studied in the context of simulation systems, and it has often been found to impact performance in human-computer systems (Diaz & Sims, 2003). Spatial ability includes many subsets, one of which is spatial orientation. Guilford (1956) defines spatial orientation as the ability to perceive and interpret the spatial relationships between things with reference to one's self. This is important to consider in research relating to navigation through virtual environments because spatial orientation has to do with how well an individual can align himself to a reference point or location (Diaz & Sims, 2003). This subset of spatial ability can be studied by means of the widely-used Guilford-Zimmerman Spatial Orientation (GZSO) measure (Diaz & Sims, 2003).

Several studies have investigated the impact of spatial orientation on task performance involving navigation through virtual environments. Moffat, Hampson, and Hatzipantelis (1998) examined spatial route learning through virtual mazes and how spatial ability was correlated with performance. Spatial ability was measured using a variety of paper-based spatial ability tests, including the GZSO measure. Moffat et al. (1998) required participants to learn two different virtual mazes by navigating through them, and the participants were given several repeated trials. Results showed that males outperformed females in maze performance (Moffet et al., 1998). Males also scored higher on the GSZO than females. In addition, Moffet and colleagues (1998) found that higher spatial ability, as measured by the GZSO, was correlated with faster navigation and fewer navigation errors.

These studies suggest that people with higher spatial orientation may have an advantage in virtual environments over those people with low spatial orientation scores. Part of the task in the current study requires navigation through a virtual environment, and therefore spatial orientation (measured using the GZSO) was used as a covariate. Participant performance was expected to be affected by spatial orientation in the current research because the participant must be aware of changes in direction and position when completing the experimental tasks.

*Video Game Experience (VGE)*

Video game experience (VGE) is also an important variable to consider when implementing simulation-based training (Orvis, Horn, & Belanich, 2009), and it may influence the effectiveness of feedback interventions in the current experiment. A previous study using the same simulation environment as used in the current research suggested that VGE is associated with performance advantages (Priest, Durlach, & Billings, in review). Participants were tasked to monitor a UAV on one computer screen while searching for targets on another computer screen, using an avatar to navigate through a virtual environment. Results indicated that individuals who reported more video game experience tended to perform better on the search tasks (Priest et al., in review).

Other existing research also supports the notion that video game experience gives individuals a performance advantage when completing tasks in simulation-based environments that have characteristics of video games. Orvis et al. (2009) conducted research that examined how several different individual differences affected performance and motivation to learn in a videogame-based training system called *America's Army*. This particular game offered a first-person perspective in a simulated small military team environment and was used to train military tactics, which is very similar to the GDIS game used in the current experiment. The researchers conducted hierarchical regression, and results indicated that VGE accounted for a significant amount of variance in all of the learner outcomes, including training performance ($R^2 = .11$, $p < .01$; Orvis et al., 2009). Another study examined how modifying task difficulty during game play affected training performance and motivation (Orvis, Horn, & Belanich, 2008). Results showed that participants with prior VGE performed better across task difficulty conditions than those participants without prior experience, regardless of mission difficulty (Orvis et al., 2008). One reason why this effect may have been evident is because people with prior VGE have the advantage of knowing more game-play strategies, using prior schema to reduce extraneous load associated with game-play, and increased memory and solution speed (Orvis et al., 2008). These things "…should all be relevant to successful performance in a fast paced, dynamic game-based instructional environment (regardless of task difficulty condition) compared to novice gamers" (Orvis et al., 2008; p. 2428). Based on empirical evidence in these prior investigations, VGE was used as a covariate in the current research.

## Experimental Hypotheses

Considering the existing literature and the research aims of this study, several testable hypotheses were created.

### *Hypothesis 1*

It is hypothesized that the four formative feedback groups (i.e., the participants who are given corrective information in addition to outcome feedback) will exhibit learning benefits from the feedback intervention, while the control group (i.e., the participants who are given only outcome feedback) will not. Hypothesis one is based on theoretical research suggesting that formative feedback which identifies how to improve performance is more effective than feedback that simply indicates the correctness of an action or answer (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kluger & DeNisi, 1996). In fact, in their meta-analysis, Bangert-Drowns and colleagues (1991) found a very low average effect associated with the use of right/wrong feedback, where learners were only given information pertaining to the correctness of their answers. Conversely, when learners were given additional information regarding the correct answer, or if they were specifically guided to the right answer, feedback was more effective (Bangert-Drowns et al., 1991). Results of this meta-analysis suggest that feedback is most effective when it includes information that in some way informs learners of the correct answer instead of using simple outcome feedback, which only indicates how correct a learner is. In addition, research on the Cognitive Load Theory (CLT) indicates that giving individuals more information can improve learning. Therefore, support for this hypothesis would provide further evidence for the CLT. There are three specific predictions relating to the first hypothesis. It is predicted that these four formative feedback groups will (1) show performance improvement over time, (2) show higher performance on the transfer mission (Mission 4) than the control group, and (2) demonstrate higher knowledge post-test scores than the control group.

*Prediction 1*

The first prediction is that all of the formative feedback groups will show performance improvement over time, while the control group will not. Performance scores across missions are expected to increase for the four formative feedback groups. Performance scores are expected to remain unchanged over time for the control group.

*Prediction 2*

The second prediction is that all of the formative feedback groups will demonstrate higher performance scores on the transfer mission (Mission 4) than the control group.

*Prediction 3*

The third prediction is that all of the formative feedback groups will score higher on the knowledge post-test than the control group.

### *Hypothesis 2*

It is also hypothesized that the bottom-up condition will be a more effective feedback intervention than the other three formative feedback groups in terms of performance, retention, and cognitive load. This hypothesis is based on the theories of cognitive load and expertise reversal effect. According to the CLT and ERE, adaptive bottom-up feedback should produce the most positive effects for novice learners. Learning improves because detailed feedback decreases the need for novices to integrate information themselves (lowering cognitive load; McLaughlin et al., 2008). Hence, performance should be better, and cognitive workload scores for this group should also be lower overall. According to the ERE, bottom-up feedback should be superior; novices should perform well with detailed feedback, and they should perform better with general feedback as their performance improves. Based on these theories, three specific predictions were made.

*Prediction 1*

The first prediction is that the bottom-up condition will show higher overall performance scores across missions than the other three formative feedback groups.

*Prediction 2*

The second prediction is that the bottom-up condition will score higher on the knowledge post-test than the other three formative feedback conditions.

*Prediction 3*

The third prediction is that the bottom-up condition will report lower cognitive load scores on the final transfer mission (Mission 4) than the other three formative feedback groups.

## Participants

An a priori power analysis was conducted using the G*Power 3 computer program (Faul et al., 2007). The following inputs were used for the power analysis: (1) medium estimated effect size of $f = .25$; (2) $\alpha = .05$; (3) desired power level = .80; (4) nonsphericity correction $\varepsilon = 1$; and (5) correlation between repeated measures = .50. The correlation estimate was a conservative estimate because nothing has been done with the repeated measures used in the current experiment, and some degree of correlation between repeated measures was likely. The inputs, results, and graphical illustrations of the required power for this study are shown in Appendix E. It was determined that 125 total participants are needed (25 per condition) to achieve a power level of .80.

Participants were primarily recruited from the UCF Psychology Department's online recruitment tool. In order to supplement this recruiting, however, word of mouth and announcements in classrooms were also used. Participants were informed that they could choose to withdraw at any time and receive credit or money for the time they spent participating. Inclusion criteria required participants to be at least 18 years old, have normal or corrected vision, and have normal manual dexterity for operating a desktop computer game. Before beginning the experiment, each participant signed an informed consent form (Appendix F).

Sixty-five males and 65 females from the University of Central Florida area volunteered to participate in this experiment in exchange for monetary compensation or college course credit. Eleven other participants did not complete the entire experiment (8 female, 3 male). An equal number of males and females were in each feedback condition to account for possible gender differences in video game experience as well as spatial orientation ability. Of the 130 participants who completed the entire experiment, the mean age was 20.39 years old, and the participants varied in their reported levels of computer skills and video game experience. Eight participants reported that they were computer novices, while 102 reported that they had intermediate computer skills. An additional 20 participants reported that they were experts on the computer. Ninety participants reported that they owned a video game system, and 40 reported that they did not own one. When asked to rate their gaming skills, 53 participants reported being novices, 57 were intermediate gamers, and 20 were expert gamers. Participant responses were also varied on several other questions regarding video game experience (See Table 2). Each participant signed an informed consent before any testing began.

Table 2.Participants' Computer and Video Game Experience

| *Participants (N = 130)* | | *Number* | *Percentage* |
|---|---|---|---|
| What is your confidence with video games? | | | |
| | Very low | 19 | 14.6 |
| | Low | 22 | 16.9 |
| | Average | 40 | 30.8 |
| | High | 25 | 19.2 |
| | Very high | 24 | 18.5 |
| How many hours per week do you play video games? | | | |
| | 0-9 hours | 98 | 75.4 |
| | 10-19 hours | 25 | 19.2 |
| | 20-29 hours | 5 | 3.8 |
| | 30-39 hours | 2 | 1.5 |
| How often do you play 1st person shooter games? | | | |
| | Never | 46 | 35.4 |
| | Rarely | 32 | 24.6 |
| | Monthly | 15 | 11.5 |
| | Weekly | 28 | 21.5 |
| | Daily | 9 | 6.9 |

## Materials

Paper-based training manuals were used to prepare participants for the tasks involved in this study. In addition, a combination of paper-based subjective measures and questionnaires were used. They included a demographics questionnaire, the Guilford-Zimmerman Spatial Orientation test, the Cognitive Load Questionnaire, a knowledge pre-test, a knowledge post-test, the Feedback Experience Questionnaire, and the Motion History Questionnaire.

### *Training Manuals*

Two paper-based training manuals were used in this experiment. The system training manual (including pictures and text) explained how to operate and navigate the video game used to simulate the search and rescue task. It also explained how to use the computerized texting and automated feedback system (TAFS). The mission protocol training manual explained the correct procedures for the missions, outlined the specific learning objectives, and informed the participant that he needed learn and apply these learning objectives to the subsequent search and rescue missions.

### *Demographics Questionnaire*

The demographics questionnaire was a paper-based measure that included questions pertaining to an individual's age, ethnicity, and computer experience. It also included several questions relating to video game experience to assess how much experience an individual had with video games in general (See Appendix G).

### *Guilford-Zimmerman Spatial Orientation Test*

The Guilford-Zimmerman Spatial Orientation (GZSO) measure was part of a larger aptitude test battery called the Guilford-Zimmerman Aptitude Survey (Guilford & Zimmerman, 1948; Guilford, 1956). The paper and pencil GZSO measure included 60 items. Each item contained two graphical images: the first one showed an image of a boat, looking over the prow towards the horizon, and landmarks in the distance. The second image showed the same boat, but the direction of the boat had changed slightly, and hence the landscape had changed slightly, relative to the first image. The participant was required to gauge how the boat had moved in heading (right, left) and position (tilting to the left or right, tilting downward or upward) in the second image, relative to the original position in the first image. Each test item included five alternative answers from which to choose, and 10 minutes were allotted to complete as many items as possible. The GZSO was scored by subtracting one-fourth of the incorrect answers from the correct answers, yielding a maximum score of 60.

### *Cognitive Load Questionnaire (CLQ)*

The Cognitive Load Questionnaire (Paas, 1992) was a one-item questionnaire that requested an individual to gauge how much mental effort he or she invested in completing a task. This questionnaire used a 9-point Likert scale, ranging from very, very low mental effort (1) to very, very high mental effort (9).

### *Knowledge Pre-Test and Post-test*

The knowledge pre-test consisted of true and false questions relating to the specific learning objectives and how to apply them to search and rescue tasks. This test measured an individual's prior knowledge of the task before he or she began training. The knowledge post-test also consisted of true and false items relating to the learning objectives and the training task. It measured the declarative knowledge of the individual after training had been completed (See Appendix H).

### *Feedback Experience Questionnaire*

The Feedback Experience Questionnaire was a measure of subjective opinions of feedback. It was created and used in Van Duyne's (2001) feedback research, and later it was adapted for Bolton's (2006) feedback research. There were eleven items on the questionnaire consisting of statements relating to how comprehensible the feedback was, how much it contributed to better performance, and how participants utilized the feedback. Participants rated each item on a 6-point Likert scale, with 1 indicating "strongly disagree" and 6 representing "strongly agree." There were an additional four items specifically for participants who received no feedback during training (or felt that they received no feedback). These four items included statements pertaining to how participants felt feedback could have helped them perform (See Appendix I).

### *Motion History Questionnaire (MHQ)*

There was little to no risk in participating in this experiment other than what would be expected during normal computer use. Even so, the Motion History Questionnaire (MHQ) was given before the study commenced in order to identify people who may be very susceptible to simulator sickness (Kennedy et al., 2001). The MHQ was a questionnaire that asked participants about various experiences they have had with motion sickness. In addition, participants were observed during the study to make sure they were not negatively affected by the simulation.

## Apparatus

### *GDIS & TAFS*

The simulation environment that the participant used included two different system components that were run on two separate desktop computers. These two computers were positioned adjacent to each other, and each computer had a 20-inch widescreen monitor. The Game Distributed Interactive Simulation (GDIS) system ran on one computer and consisted of the video game in which participants performed the simulated missions. The texting and automated feedback system (TAFS) ran on the second computer and was used to communicate during missions, as well as to deliver semi-automated feedback to participants upon completion of the training missions. A confederate used a third computer to run the experimenter's version of TAFS so that feedback could be sent to the participants.

The GDIS system was a virtual immersive environment similar to a first-person shooter style video game. GDIS has capabilities to include semi-intelligent computer generated forces during game play and can also support distributed team training; however for this study no weapons or other human players were present in the simulation. GDIS simulated a realistic MOUT (Military Operations in Urban Terrain) site located in Ft. Benning (See Figure 2 for an example of the interface). Participants used a keyboard and a standard two-button, one-wheel Dell optical mouse to control and navigate their avatars through the simulated environment. Participants were able to traverse the environment, open and close doors, and explore the interiors of buildings.


Figure 2. Sample of the GDIS Environment.

While participants controlled an avatar in GDIS, they also simultaneously operated TAFS using a keyboard and a standard two-button, one-wheel Dell optical mouse. TAFS had three functions: (1) it allowed participants to communicate with Headquarters (HQ; a confederate played this role), (2) it allowed the confederate to monitor and input all participant protocol-related actions so that the system could generate the appropriate feedback, and (3) it displayed feedback for the participant to peruse after each training mission. Participants were able to type text communications and send them to HQ, and they were also able to view messages sent from HQ. In the TAFS experimenter's interface, the confederate was able to monitor and log all participant actions (including actions and inactions; See Figure 3 for an example of the experimenter's interface). Using these inputs, TAFS automatically generated the appropriate feedback for participants on the basis of their feedback condition. This feedback was displayed on the participant's computer screen and remained there until the participant clicked a button to close the feedback notification window. See Figure 4 for an example of a feedback notification window. Text logs were automatically saved to a file for data collection purposes. Data collection occurred in a standard laboratory environment at the Institute for Simulation and Training, located in Orlando, Florida.

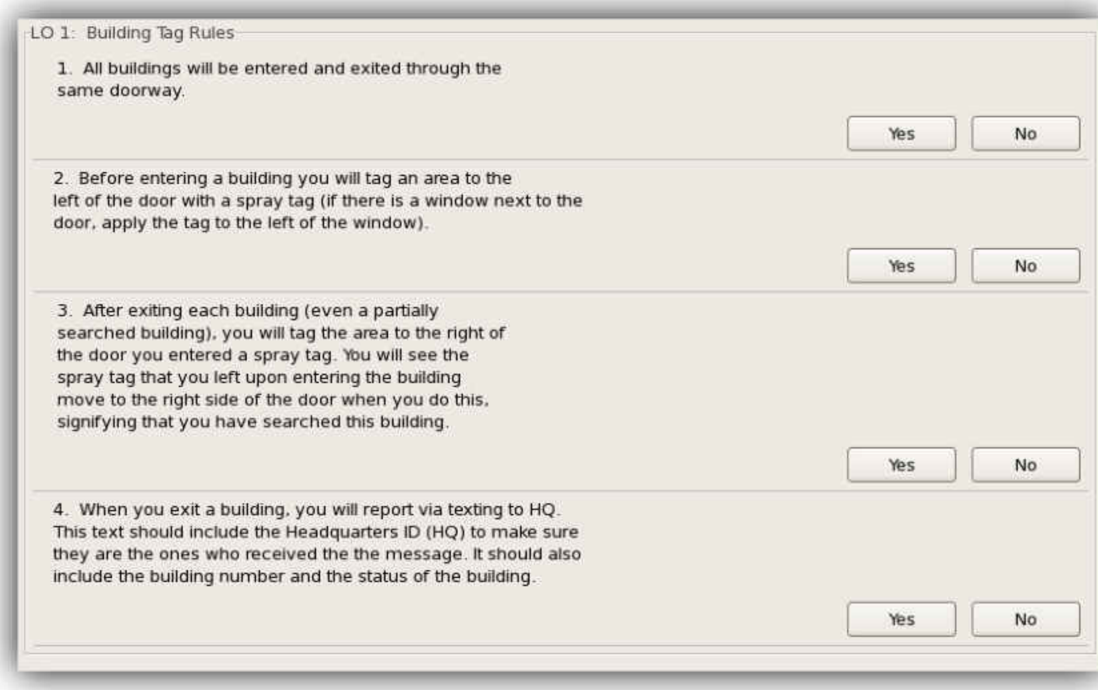



Figure 3. Example of the Experimenter Interface for TAFS. Participant Actions and Inactions Were Logged by Clicking the "Yes" or "No" Buttons.

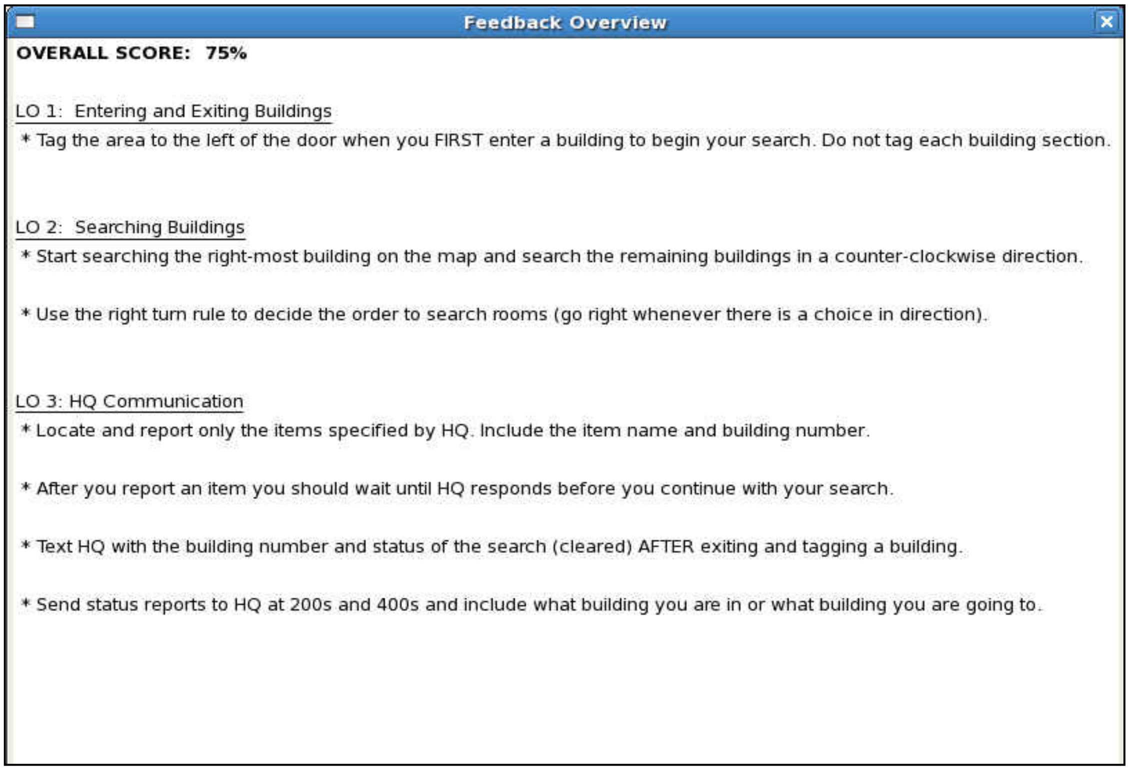



Figure 4. Example of a Feedback Notification Window Displayed on the Participant's Computer Screen.

## Performance Measures

### *Overall Performance Scores*

Paper-and-pencil knowledge tests were given to measure declarative and procedural knowledge, but individuals also needed to show their ability to apply this knowledge in a simulated environment in different scenarios. Therefore, measures of performance were included so that individuals could demonstrate that they could also integrate information. Overall performance scores were calculated for each of the four missions, and these were the primary dependent measures for data analysis. These percentage scores were calculated based on the number of actions performed correctly, out of the total number of opportunities to perform a correct action for each learning objective during a mission. Participants were given 10 minutes to complete each mission, and many participants did not complete the entire search within this allotted time. Consequently, participants were only scored on actions that they performed correctly (or had the opportunity to perform correctly) within those 10 minutes. In this way, participants were not penalized for failing to complete the mission within the time limit. The overall performance scores were then examined to see if performance declined or improved over missions, as a result of feedback condition. Performance improvement is a critical criterion to measure because "…an individual's performance while completing a training program is indicative of the extent to which he/she is acquiring the knowledge and skills being targeted within the instructional content" (Orvis et al., 2008, p. 2416).

### *Pre-Test and Post-Test Knowledge Scores*

A written knowledge pre-test was administered before training to make sure that participants were equally knowledgeable about the subject domain. It was assumed that the participants would not differ significantly in their scores on the pre-test. A written knowledge post-test was administered after the transfer mission to measure how much declarative knowledge the participant had learned during training.

### *Cognitive Load Questionnaire (CLQ)*

The CLQ was a subjective measure of cognitive load and was administered following each mission. In addition, a baseline measure was collected before training missions commenced.

## Procedure

Participants were first required to read and sign an informed consent (See Appendix F), and they were given information about their privacy in accordance with the Privacy Act. Participants were randomly assigned to one of the five conditions: (1) detailed, (2) general, (3) bottom-up, (4) top-down, or (5) the control group. They completed the MHQ, the demographics questionnaire, the knowledge pre-test, and the Guilford-Zimmerman Spatial Orientation measure.

Participants were given a systems training manual where they learned how to operate the GDIS game (i.e. how to navigate, perform actions, etc.) as well as how to operate TAFS. After participants read through the systems training manual, they were asked to perform a simple task in a virtual environment similar to the one used in the training missions in order to demonstrate navigation and GDIS game function proficiency. Participants were not given a performance score for this practice exercise. Once they successfully completed this practice exercise, participants completed the baseline measure of the Cognitive Load Questionnaire (CLQ). Following this, participants were given the mission training manual, where they learned about the three learning objectives and search and rescue mission protocol. Participants were given enough time to review the training manual until they felt comfortable with the information. They were given the opportunity to ask any relevant questions at this time.

After initial training was completed and the necessary baseline measures were established, participants completed three different training missions and one transfer mission (Mission 4). The transfer mission was included because true comprehension of concepts can be best reflected in an individual's ability to apply newly acquired information in new situations (Moreno, 2004). Hence, we are most likely to correctly determine whether learning took place by examining performance in a new or different situation. In this particular study, the transfer mission included new visual cues and slightly modified mission requirements.

For each training mission (Missions 1, 2, and 3), the procedure was the same. First, the participant received a mission briefing that outlined the mission and assigned a set of buildings to search. Second, the participant began the training mission and had 10 minutes to complete as much of the search as he or she could. Once the time limit expired the mission ended, and participants then completed the CLQ. Then participants received feedback for that mission, based on their respective feedback condition. The transfer mission (Mission 4) followed the same procedures, except no feedback was given. After the transfer mission, participants completed the CLQ and a question asking them how motivated they were to perform well on the missions. Next, they completed the feedback experience questionnaire and the knowledge post-test. Finally, participants were debriefed on the nature of the experiment and were compensated for their participation.

# CHAPTER FOUR: RESULTS

## Data Analysis Plan

Analyses were performed using SPSS 14.0 and Statistica 7 for Windows. An alpha level of .05 was used for all analyses, unless otherwise noted. Before any analyses were performed, the data were examined for any issues that could potentially affect the results of the statistical analyses.

First, a manipulation check was performed by examining data for each learning objective (LO) for each participant in the two adaptive feedback groups to make sure that all participants actually received adaptive feedback for the three learning objectives (i.e., feedback content switched between detailed and general at some point across the missions). The findings of this manipulation check indicate that every participant did not experience a change in feedback for all three learning objectives (See Appendix J for a thorough explanation of the procedure). However, it is still possible that each individual experienced adaptive feedback at some point during the training session as a whole, with changes occurring for some learning objectives but perhaps not for others. Therefore, the entire training session was examined. Findings confirmed that every participant in the bottom-up and top-down feedback conditions experienced the manipulation during training, although not every participant experienced it for all three learning objectives (See Appendix J for details).

Second, normality was checked for all dependent variables, using the Kolmogorov-Smirnov (K-S) test for normality. There were several instances where the dependent variables in the feedback groups were not normally distributed, which is quite common in larger samples according to Pallant (2007; See Table 3). The existence of several outliers in the raw performance data was also noted. These outliers were included in the analyses in order to avoid any bias in transforming or discarding the outlying scores. The participants all came from the same population; therefore the high and low scores were most likely due to chance and should be included in the statistical analyses, especially since the primary dependent variables were performance scores. Rather than risk changing the distribution entirely by not taking into account extreme scores, the true distributions of the performance scores were maintained for the analyses.

Table 3. Violations of Normality for Performance Scores, Knowledge Post-test Scores, and Cognitive Load.

| *Feedback Condition* | *Dependent Variable* | *K-S Statistic* | *df* | *Sig.* |
|---|---|---|---|---|
| Constant Detailed | Mission 2 score | .206 | 26 | .006 |
| | Mission 4 score | .197 | 26 | .011 |
| | Post-test score | .198 | 26 | .010 |
| | Cognitive Load (Mission 4) | .198 | 26 | .010 |
| Constant General | Post-test score | .188 | 26 | .019 |
| | Cognitive Load (Mission 4) | .223 | 26 | .002 |
| Adaptive Bottom-up | Post-test score | .260 | 26 | <.001 |
| | Cognitive Load (Mission 4) | .189 | 26 | .018 |
| Adaptive Top-down | Post-test score | .261 | 26 | <.001 |
| | Cognitive Load (Mission 4) | .245 | 26 | <.001 |
| Control | Cognitive Load (Mission 4) | .216 | 26 | .003 |

According to Hays (1994), sometimes it is necessary to run statistical analyses (in particular, ANOVA and the *F* test) when some of the assumptions (including normality) have not been met. Hays (1994) points out that, "It can be shown that, other things being equal, inferences made about means that are valid for normal populations also are valid even when the forms of the population distributions depart considerably from normal, provided that the *n* in each sample is relatively large" (p. 406). Pallant (2007) also agrees that, "…with large enough sample sizes, the violation of this assumption [normality] should not cause any major problems" (p. 204). Therefore, the analyses pertaining to overall performance scores for missions were conducted using ANCOVA. However, knowledge post-test scores and cognitive load scores had extensive violations of normality, so non-parametric Kruskal-Wallis analyses were used for these data.

Following these preliminaries, the covariates (CVs; video game experience and spatial orientation) were examined. The responses to four demographic questions (where participants rated their own experiences with video games) were found to be significantly correlated (See Table 4). Therefore, these variables were standardized and combined into a single measure of video game experience (VGE). The five feedback conditions were then checked for equivalence on VGE and spatial orientation ability, and no significant differences across feedback conditions were detected.

Table 4. Spearman Intercorrelations between Demographics Questions Related to Video Game Experience.

| *Demographics Question* | *1* | *2* | *3* | *4* |
|---|---|---|---|---|
| | Participants ($N$ = 130) | | | |
| 1. How would you rate your video game skills? | --- | .852** | .478** | .675** |
| 2. What is your level of confidence with video games? | | --- | .566** | .748** |
| 3. How many hours/week do you play video games? | | | --- | .588** |
| 4. How often do you play first-person shooters? | | | | --- |

** correlation is significant at the .01 level (2 tailed)

Independent samples t-tests were conducted to compare VGE for males and females, as well as spatial orientation for males and females. For VGE, Levene's test was significant, $F$ = 9.826, $p$ = .002, so equal variances were not assumed. Males ($M$ = 2.363, $SD$ = 2.903) reported significantly higher video game experience than females ($M$ = -2.363, $SD$ = 1.872), $t$ (109.367) = 11.030, $p$ < .001. The magnitude of the differences in the means (mean difference = 4.725, 95% CI: 3.876 to 5.574) was very large ($\eta^2$ = .487). On the Guilford-Zimmerman Spatial Orientation Test, males ($M$ = 22.162, $SD$ = 11.980) scored significantly higher than females ($M$ = 15.119, $SD$ = 9.886), $t$ (128) = 3.656, $p$ < .001 (two-tailed). The magnitude of the differences in the means (mean difference = 7.042, 95% CI: 3.230 to 10.854) was fairly large ($\eta^2$ = .095).

The CVs were examined further to make sure that all assumptions for covariates were met before running any ANCOVA analyses. Both VGE and spatial orientation were continuous variables, and they were significantly correlated with each other, but not too strongly (Pearson $r$ = .298). Spatial orientation and VGE had linear relationships and were also correlated significantly with the overall performance scores for the missions (the dependent variables). Finally, the assumption of homogeneity of regression was met for these CVs, and there were no interactions between the CVs and the experimental manipulation. Therefore, both CVs were included in the statistical analyses.

Mixed-model and one-way ANCOVAs were used to analyze the mission performance data, and VGE and spatial orientation were included as CVs. In addition, the non-parametric Kruskal-Wallis analysis was used to examine knowledge post-test scores and cognitive load scores. Lastly, exploratory analyses were conducted on the Feedback Experience Questionnaire and spatial orientation scores.

## Hypothesis 1

The first hypothesis suggested that the four formative feedback groups would exhibit learning benefits from the feedback intervention, while the control group (i.e., those who received only outcome feedback) would not. Three specific predictions were addressed regarding learning, performance, and retention of information.

### *Prediction 1*

For hypothesis one, the first prediction was that all formative feedback groups would show performance improvement over time, while the control group would not. A mixed between-within subjects ANCOVA was conducted to assess the impact of five different feedback interventions (detailed, general, bottom-up, top-down, and control group) on participants' overall performance scores across missions (Mission 1, Mission 2, Mission 3, and Mission 4). VGE and spatial orientation were used as CVs. Table 5 presents the mean overall performance scores and standard deviations for each of the feedback conditions during the four missions; Figure 5 shows the mean performance scores graphically. The analyses revealed a significant main effect of mission ($F(3, 369) = 5.292$, $p = .001$, $\eta_p^2 = .041$, power = .929), a significant mission-by-condition interaction ($F(12, 369) = 4.562$, $p < .001$, $\eta_p^2 = .129$, power = 1.000), and a significant main effect of feedback condition ($F(4, 123) = 12.221$, $p < .001$, $\eta_p^2 = .284$, power = 1.00). Participants tended to improve their performance over missions and to perform differently according to which type of feedback intervention they received. Improvements in performance also differed based on feedback intervention. In addition, spatial orientation was a significant CV, $F(1, 123) = 26.289$, $p < .001$, $\eta_p^2 = .176$, power = .999. Higher spatial orientation scores were associated with higher performance across missions (for Mission 1, Pearson $r = .381$; for Mission 2, Pearson $r = .384$; for Mission 3, Pearson $r = .324$; for Mission 4, Pearson $r = .397$).

Table 5. Mean Overall Performance Scores and Standard Deviations for Each Feedback Condition Across the Four Missions, and Knowledge Pre- and Post-test Scores.

| Feedback Condition | | Training Missions | | | Transfer Mission | Knowledge Pre-test | Knowledge Post-Test |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | | |
| Constant Detailed (N = 26) | *M* | 71.269 | 71.962 | 77.154 | 84.846 | 71.923 | 80.000 |
| | *SD* | 14.668 | 15.795 | 12.905 | 10.631 | 14.702 | 13.267 |
| Constant General (N = 26) | *M* | 67.500 | 66.115 | 67.038 | 73.308 | 70.385 | 66.923 |
| | *SD* | 10.187 | 10.749 | 12.492 | 9.333 | 12.484 | 16.916 |
| Adaptive Bottom-up (N = 26) | *M* | 70.654 | 78.808 | 82.192 | 85.269 | 70.000 | 81.154 |
| | *SD* | 11.682 | 12.332 | 11.682 | 7.816 | 11.314 | 15.054 |
| Adaptive Top-down (N = 26) | *M* | 69.154 | 65.692 | 69.154 | 80.231 | 66.923 | 78.846 |
| | *SD* | 10.869 | 14.639 | 10.869 | 14.075 | 12.254 | 15.054 |
| Control (N = 26) | *M* | 62.539 | 61.154 | 62.539 | 61.846 | 70.000 | 61.154 |
| | *SD* | 13.923 | 13.160 | 13.923 | 18.017 | 10.198 | 20.460 |
| Overall (N = 130) | *M* | 68.223 | 68.746 | 72.031 | 77.100 | 69.846 | 73.615 |
| | *SD* | 12.593 | 14.584 | 14.764 | 15.141 | 12.199 | 18.001 |

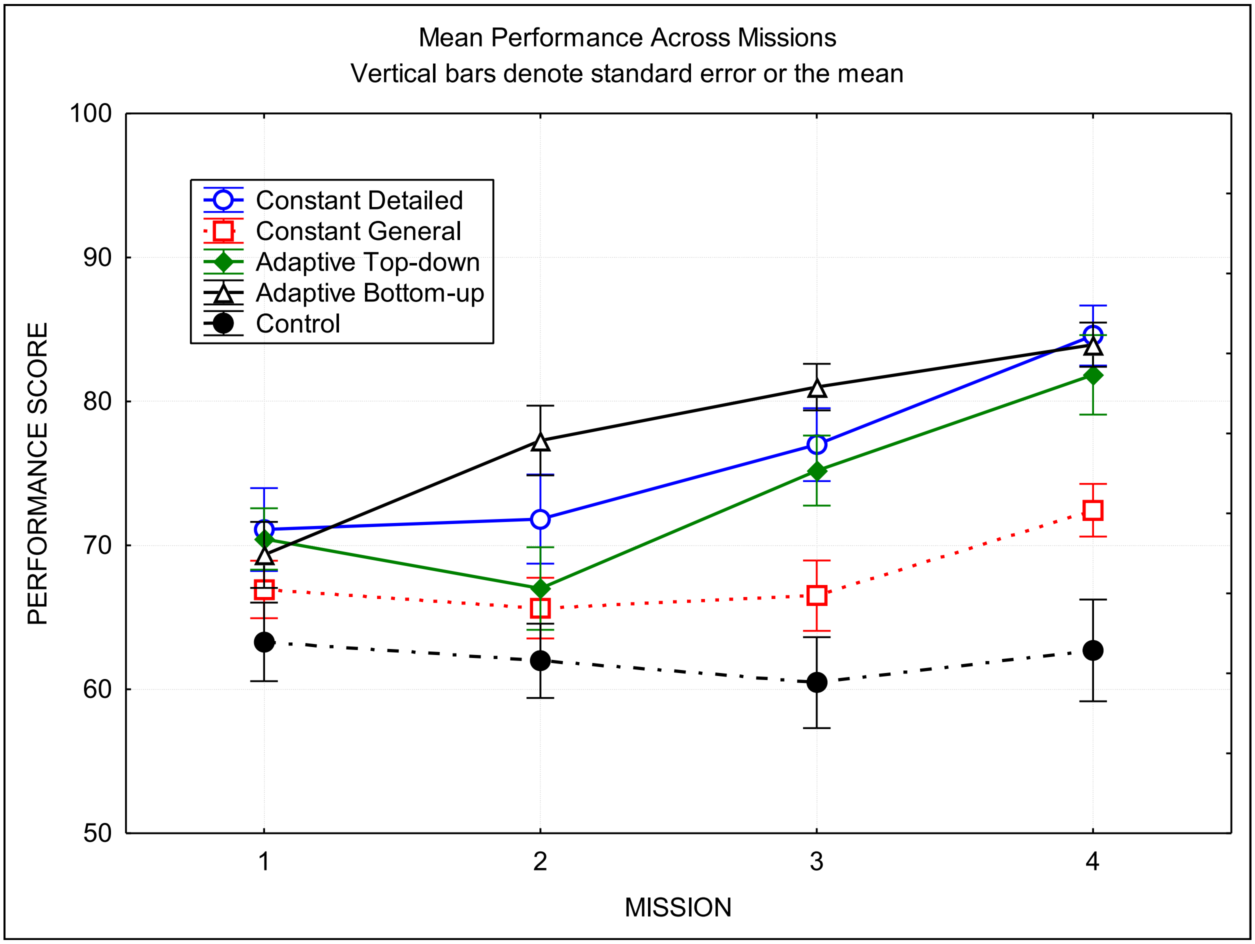



Figure 5. Average Overall Performance Scores per Mission by Feedback Condition.

Because there was an interaction, a separate repeated measures ANCOVA was performed for each feedback condition to see if a change in performance occurred across missions. Results for the detailed condition showed a significant effect of mission, $F(3, 69) = 3.135$, $p = .031$, $\eta_p^2 = .120$, power = .705, where participants' performance increased over the missions. A significant effect of spatial orientation was also found, $F(1, 23) = 20.072$, $p < .001$, $\eta_p^2 = .466$, power = .990, where higher spatial orientation scores were associated with higher performance on

Mission 1 (Pearson $r = .535$), Mission 2 (Pearson $r = .681$), Mission 3 (Pearson $r = .676$) and Mission 4 (Pearson $r = .688$).

No effect of mission was found for the general condition, $F(3, 69) = .344$, $p = .793$, $\eta_p^2 = .015$, power = .114. However, there was a significant effect of spatial orientation, $F(1, 23) = 5.100$, $p = .034$, $\eta_p^2 = .181$, power = .581. For participants in this condition, higher spatial orientation scores were associated with higher performance on Mission 1 (Pearson $r = .305$), Mission 2 (Pearson $r = .265$), Mission 3 (Pearson $r = .232$) and Mission 4 (Pearson $r = .397$).

Results for the bottom-up condition revealed a significant effect of mission, $F(3, 69) = 4.155$, $p = .009$, $\eta_p^2 = .153$, power = .833, showing a performance increase over missions. There was also a significant effect of VGE, $F(1, 23) = 11.108$, $p = .003$, $\eta_p^2 = .326$, power = .891. For participants in this condition, more video game experience was associated with higher performance in Mission 1 (Pearson $r = .345$), Mission 2 (Pearson $r = .434$), Mission 3 (Pearson $r = .196$), and Mission 4 (Pearson $r = .557$).

The top-down condition results also showed a significant effect of mission, $F(3, 69) = 2.828$, $p = .045$, $\eta_p^2 = .110$, power = .655, showing that participants' performance increased as they went through the missions. A significant effect of spatial orientation was also found, $F(1, 23) = 7.117$, $p = .014$, $\eta_p^2 = .236$, power = .724. For these participants, higher spatial orientation scores were associated with higher performance scores for Mission 1 (Pearson $r = .333$), Mission 2 (Pearson $r = .349$), Mission 3 (Pearson $r = .552$), and Mission 4 (Pearson $r = .491$).

Finally, no effect of mission was found for the control group, $F(3, 69) = 0.543$, $p = .654$, $\eta_p^2 = .023$, power = .156, although there was a significant effect of spatial orientation, $F(1, 23) = 4.734$, $p = .040$, $\eta_p^2 = .171$, power = .550. For this group, higher spatial orientation scores were associated with higher performance scores on Mission 1 (Pearson $r = .453$), Mission 2 (Pearson $r = .346$), Mission 3 (Pearson $r = .259$), and Mission 4 (Pearson $r = .443$).

In summary, the detailed, bottom-up, and top-down feedback interventions showed significant improvement in performance over time. The general group and the control group did not demonstrate any significant change in performance over missions. Although the prediction that all the formative feedback groups would demonstrate performance improvement over missions was not supported, the pattern suggests that detailed feedback is important for improvement; the two groups who did not show improvement were the groups that did not receive detailed feedback at any point during training (constant general and control groups).

### *Prediction 2*

The second prediction for hypothesis one was that all formative feedback groups would demonstrate better performance than the control group on the transfer mission (Mission 4). A one-way between-groups ANCOVA was conducted to compare the effectiveness of four formative feedback interventions and one control group. The independent variable was the type of feedback intervention (detailed, general, bottom-up, top-down, and control group), and the dependent variable consisted of overall performance scores for Mission 4. Participants' spatial orientation scores and VGE were used as the covariates in this analysis. See Table 5 for means and standards deviations of performance scores on Mission 4; Figure 6 shows the mean performance scores for Mission 4 graphically. Results showed significant performance differences between feedback interventions, $F(4, 123) = 18.010$, $p < .001$, $\eta_p^2 = .369$, power = 1.000. There was also a significant effect of spatial orientation, $F(1, 123) = 26.816$, $p < .001$, $\eta_p^2$

= .179, power = .999, with higher spatial orientation scores associated with higher performance scores on Mission 4 (Pearson $r$ = .397).

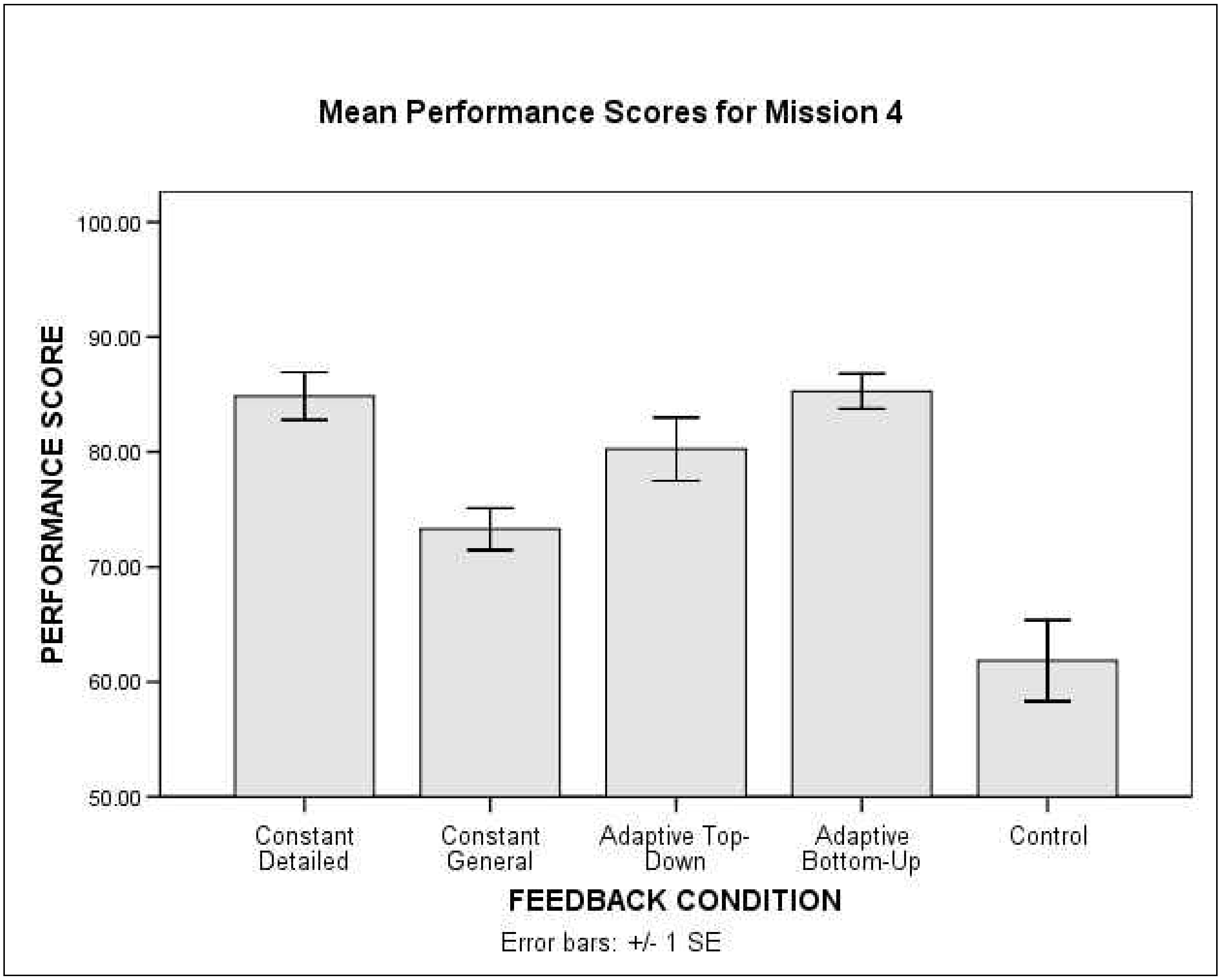



Figure 6. Mean Performance Scores on Mission 4 for Feedback Conditions.

Planned comparisons between each of the formative groups and the control group were conducted, and the results are summarized in Table 6. Participants in all formative feedback groups performed at significantly higher levels than participants in the control group. The prediction that all formative groups would demonstrate better performance on Mission 4 was supported.

Table 6. Planned Comparisons between Each Formative Feedback Group Versus the Control Group for Performance Scores on Mission 4.

| Condition | *t* | *p* |
|---|---|---|
| Detailed vs. control | (50) = 6.982 | <.001 |
| General vs. control | (50) = 3.096 | .002 |
| Bottom-up vs. control | (50) = 6.726 | <.001 |
| Top-down vs. control | (50) = 6.112 | <.001 |

### *Prediction 3*

The third prediction for hypothesis one was that all formative feedback groups would score higher on the knowledge post-test than the control group. It was assumed that participants would not differ significantly between groups on the knowledge pre-test, which was given before training commenced. See Table 5 for means and standard deviations of the pre-test scores. Therefore, first a Kruskal-Wallis analysis was run to evaluate differences between groups on knowledge pre-test scores, and no differences between feedback condition were found, $\chi^2(4, N = 130) = 2.565$, $p = .633$. As expected, participants in the feedback groups did not differ significantly in terms of knowledge pre-test scores. The knowledge post-test scores were examined next.

A Kruskal-Wallis test was conducted to evaluate differences among the five feedback conditions (detailed, general, bottom-up, top-down, and control) on median change in knowledge post-test scores (See Table 5 for means and standard deviations of knowledge post-test scores;

Figure 7 shows the mean knowledge post-test scores graphically). Significant differences between feedback interventions were found, $\chi^2(4, N = 130) = 24.726, p < .001$, so the Mann-Whitney *U* test was used to evaluate planned pairwise comparisons among the formative feedback groups and the control group (See Table 7). The results of these tests indicated a significant difference between the detailed group and the control group, between the bottom-up group and the control group, and between the top-down group and the control group. The general group did not differ from the control group. While the prediction that all formative groups would score significantly higher on the knowledge post-test was not supported, the pattern suggests that detailed feedback is important for high knowledge retention; the two groups that scored the lowest (constant general and the control groups) did not receive detailed feedback at any point during training.

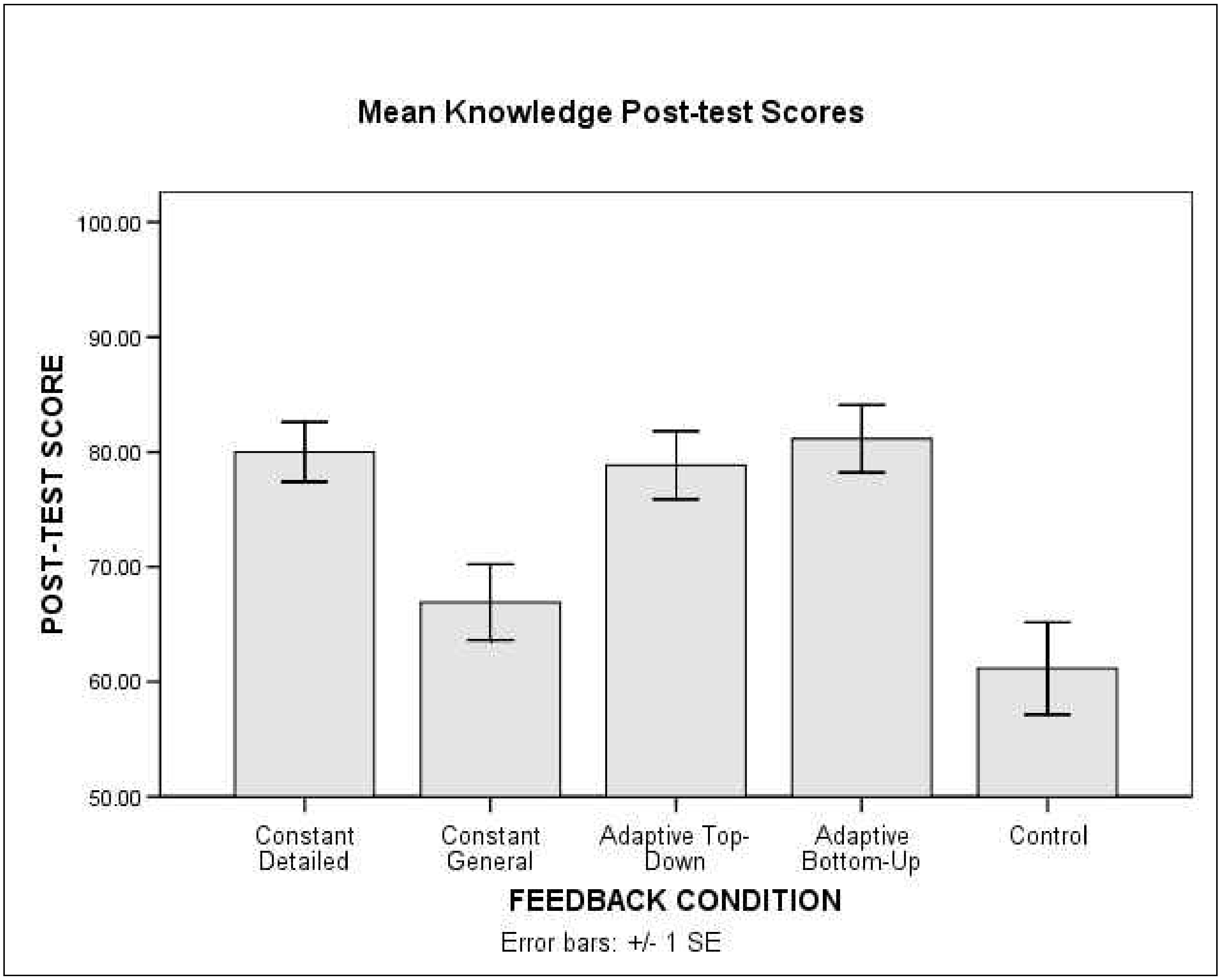



Figure 7. Average Knowledge Post-test Scores for Feedback Conditions.

Table 7. Mann-Whitney *U* test: Planned Comparisons between Each Formative Feedback Group Versus the Control Group for Knowledge Post-test Scores.

| Condition | *N* | Mann-Whitney *U* | Wilcoxon *W* | *z* | *p* |
|---|---|---|---|---|---|
| Detailed vs. control | 52 | 153.500 | 504.500 | -3.420 | .001 |
| General vs. control | 52 | 280.500 | 631.500 | -1.069 | .285 |
| Bottom-up vs. control | 52 | 148.000 | 499.000 | -3.524 | <.001 |
| Top-down vs. control | 52 | 159.500 | 510.500 | -3.322 | .001 |

## Hypothesis 2

The second hypothesis suggested that adaptive bottom-up feedback would be a more effective feedback intervention than the other formative feedback groups in terms of performance, knowledge retention, and cognitive load. Three specific predictions relating to this hypothesis were examined.

### *Prediction 1*

For hypothesis two, the first prediction was that the bottom-up condition would show higher overall performance scores across missions than the other formative feedback conditions. A mixed between-within subjects ANCOVA was conducted to assess the impact of four different feedback interventions (detailed, general, bottom-up, and top-down) on participants' performance scores across missions (Mission 1, Mission 2, Mission 3, and Mission 4). See Table 5 for means and standards deviations of performance scores across missions; Figure 5 depicts these means graphically.

This analysis showed a significant main effect for mission ($F(3, 294) = 7.488$, $p < .001$, $\eta_p^2 = .071$, power = .986), a significant mission-by-condition interaction ($F(9, 294) = 2.796$, $p = .004$, $\eta_p^2 = .079$, power = .958), and a significant main effect of feedback condition ($F(3, 98) = 6.906$, $p < .001$, $\eta_p^2 = .175$, power = .974). Participants in these formative feedback conditions showed a change in performance over missions and performed differently according to which feedback they received. Participant performance improvement also differed based on feedback condition. Spatial orientation and VGE were found to be significant covariates, $F(1, 98) = 20.262$, $p < .001$, $\eta_p^2 = .171$, power = .994, and $F(1, 98) = 4.889$, $p = .029$, $\eta_p^2 = .048$, power =

.591. Higher spatial orientation was associated with higher performance scores on Mission 1 (Pearson $r$ = .352), Mission 2 (Pearson $r$ = .387), Mission 3 (Pearson $r$ = .345), and Mission 4 (Pearson $r$ = .416). Higher VGE was associated with better performance on the missions as well (Mission 1, Pearson $r$ = .268; Mission 2, Pearson $r$ = .330; Mission 3, Pearson $r$ = .312; and Mission 4, Pearson $r$ = .283).

Results from planned comparisons between each formative feedback group and the bottom-up group (across missions) are summarized in Table 8. Participants in the bottom-up condition performed at significantly higher levels than those in the general condition ($F(1, 48) = 21.640$, $p < .001$, $\eta_p^2 = .311$, power = .995) and the top-down condition ($F(1, 48) = 4.492$, $p = .039$, $\eta_p^2 = .086$, power = .547). Furthermore, participants in the detailed group did not differ significantly from those in the bottom-up condition. These analyses also showed a significant mission-by-condition interaction for the general condition ($F(2.595, 124.577) = 5.251$, $p = .003$, $\eta_p^2 = .099$, power = .892) and the top-down condition ($F(3, 144) = 3.772$, $p = .012$, $\eta_p^2 = .073$, power = .804) when each was compared with the bottom-up group.

Table 8. Planned Comparisons between Each Formative Feedback Group Versus the Bottom-up Group.

| Feedback Condition | *F* | *p* | $\eta_p^2$ | power |
|---|---|---|---|---|
| **Detailed vs. Bottom-up** | | | | |
| Mission effect | (3, 144) = 6.177 | .001 | .114 | .959 |
| Mission*Condition | (3, 144) = 1.873 | .137, *ns* | .038 | .478 |
| Condition | (1, 48) = .443 | .509, *ns* | .009 | .100 |
| VGE | (1, 48) = 8.594 | .005 | .152 | .819 |
| GZSO | (1, 48) = 11.175 | .002 | .189 | .906 |
| **General vs. Bottom-up** | | | | |
| Mission effect | (2.595, 124.577) = 3.451* | .024 | .067 | .718 |
| Mission*Condition | (2.595, 124.577) = 5.251* | .003 | .099 | .892 |
| Condition | (1, 48) = 21.640 | <.001 | .311 | .995 |
| VGE | (1, 48) = 1.426 | .238, *ns* | .029 | .216 |
| GZSO | (1, 48) = 1.782 | .188, *ns* | .036 | .258 |
| **Top-down vs. Bottom-up** | | | | |
| Mission effect | (3, 144) = 4.198 | .007 | .080 | .849 |
| Mission*Condition | (3, 144) = 3.772 | .012 | .073 | .804 |
| Condition | (1, 48) = 4.492 | .039 | .086 | .547 |
| VGE | (1, 48) = 8.848 | .005 | .156 | .830 |
| GZSO | (1, 48) = 4.928 | .031 | .093 | .585 |

* The assumption of sphericity was violated; df were adjusted using Greehouse-Geisser estimates.

To further analyze these two mission-by-condition interactions, separate analyses were run on the data: one analysis used the three practice missions as a repeated measures variable, and one analysis examined performance on the final transfer mission. These analyses revealed that participants in the general group differed from the performance of those in the bottom-up condition during the training missions ($F(1, 48) = 15.773$, $p < .001$, $\eta_p^2 = .247$, power = .973) as well as the final transfer mission ($F(1, 48) = 23.970$, $p < .001$, $\eta_p^2 = .333$, power = .998). There was also a training mission-by-condition interaction, ($F(1.767, 84.819) = 7.163$, $p = .001$, $\eta_p^2 = .130$, power = .900; Mauchly's test indicated that the assumption of sphericity had been violated and therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity); therefore, missions were examined individually. Per mission analyses of the three training missions revealed that differences occurred in Mission 2 and Mission 3, but not in Mission 1 (See Table 9).

In addition, participants in the top-down condition differed from those in the bottom-up group during the training missions ($F(1, 48) = 5.270$, $p = .026$, $\eta_p^2 = .099$, power = .614), but not on the final transfer mission ($F(1, 48) = 0.583$, $p = .449$, $\eta_p^2 = .012$, power = .116).There was a training mission-by-condition interaction, ($F(2, 96) = 4.859$, $p = .010$, $\eta_p^2 = .092$, power = .790); therefore, missions were examined individually. Per mission analyses of the training missions revealed that differences occurred in Mission 2 and Mission 3, but not in Mission 1 (See Table 9).

Table 9. Planned Comparisons: One-way ANCOVA per Mission.

| Condition | Training Missions | | | Transfer Mission |
|---|---|---|---|---|
| | Mission 1 | Mission 2 | Mission 3 | Mission 4 |
| Detailed vs. Bottom-up | | | | |
| *F*(1, 48) | 0.436 | 1.944 | 1.839 | 0.159 |
| General vs. Bottom-up | | | | |
| *F*(1, 48) | 0.800 | 13.174** | 23.704** | 23.970** |
| Top-down vs. Bottom-up | | | | |
| *F*(1, 48) | 0.031 | 8.667** | 4.845* | 0.583 |

*p < .05
**p < .01

In summary, the bottom-up condition did not differ from any of the formative feedback groups in performance on Mission 1. This finding was anticipated because all participants performed the first mission immediately after reading through the training manual. Therefore, they should all have started at the same performance level on Mission 1, and this was confirmed for the formative feedback conditions.

For the other missions, the planned comparisons revealed several things. First, the detailed and the bottom-up feedback conditions performed comparably. In addition, while no significant differences were found between these conditions, the trend indicates that the bottom-up feedback promoted faster learning. Second, the general feedback condition performed significantly worse than the bottom-up group on Missions 2, 3, and 4. This suggests that the adaptive bottom-up feedback intervention is more beneficial than the constant general intervention in terms of performance. Third, the top-down condition performed significantly worse than the bottom-up condition on Missions 2 and 3, although these groups performed at the same level on Mission 4. Based on these results, the prediction that the adaptive bottom-up feedback intervention would be better than the other formative groups was not supported. Learning gains appeared to be more rapid in both the bottom-up and constant detailed groups—the two groups that received detailed feedback at the beginning on training.

### *Prediction 2*

For hypothesis two, the second prediction was that individuals receiving the adaptive bottom-up feedback intervention would score higher on the knowledge post-test than the other formative feedback conditions. A Kruskal-Wallis test was conducted to evaluate differences among the four formative feedback conditions (detailed, general, bottom-up, and top-down) on median change in knowledge post-test scores (See Table 5 for means and standard deviations of knowledge post-test scores). The test was significant, $\chi^2(3, N = 104) = 12.715$, $p = .005$, so follow-up tests were conducted using the Mann-Whitney *U* test to evaluate planned pairwise comparisons among the formative feedback groups and the bottom-up group (See Table 10). The results of these tests indicated a significant difference in knowledge post-test scores only between the bottom-up and general feedback groups. Consequently, the prediction that people receiving adaptive bottom-up feedback would perform better than the other formative groups on the knowledge post-test was not supported. Based on these results, the prediction that the adaptive bottom-up feedback intervention would score higher on the knowledge post-test than the other formative feedback groups was not supported. However, the only formative feedback condition to score significantly lower than the bottom-up group was the one that did not receive any detailed feedback during training (the constant general group).

Table 10. Mann-Whitney *U* test: Planned Comparisons between Each Formative Feedback Group Versus the Bottom-up Feedback Group for Knowledge Post-test Scores.

| Condition | *N* | Mann-Whitney *U* | Wilcoxon *W* | *z* | *p* |
|---|---|---|---|---|---|
| Detailed vs. Bottom-up | 52 | 308.500 | 659.500 | -.559 | .576 |
| General vs. Bottom-up | 52 | 175.000 | 526.000 | -3.046 | .002 |
| Top-down vs. Bottom-up | 52 | 295.000 | 646.000 | -.811 | .417 |

*Prediction 3*

The third prediction under hypothesis two was that participants receiving adaptive bottom-up feedback would report lower cognitive workload scores on the transfer mission (Mission 4) than the other formative feedback groups. A Kruskal-Wallis test was conducted to evaluate differences among the four formative feedback conditions (detailed, general, bottom-up, and top-down) on median change in cognitive load scores (See Table 11 for means and standard deviations of cognitive load scores). The test was not significant, $\chi^2(3, N = 104) = 0.496$, $p = .920$. The results of this test indicate that there are no significant differences between the Mission 4 cognitive load scores of participants for the four formative feedback conditions, and hence, the prediction was not supported. To explore further, cognitive load for the other training missions were examined using the Kruskal-Wallis test to determine if differences between feedback conditions existed. No significant differences were found for cognitive load scores for any of the missions. Although the prediction was not supported, this could be due to lack of sensitivity of the Cognitive Load Questionnaire or due to the lack of cognitive demands on the task itself.

Table 11. Mean Overall Cognitive Load Scores and Standard Deviations for Each Feedback Condition Across the Four Missions.

| Feedback Condition | | Cognitive Load Scores | | | |
|---|---|---|---|---|---|
| | | Mission 1 | Mission 2 | Mission 3 | Mission 4 |
| Constant Detailed (N = 26) | *M* | 5.08 | 5.50 | 5.42 | 5.62 |
| | *SD* | 1.79 | 1.79 | 1.70 | 1.84 |
| Constant General (N = 26) | *M* | 5.65 | 5.69 | 5.69 | 5.85 |
| | *SD* | 0.98 | 1.12 | 1.26 | 1.32 |
| Adaptive Bottom-up (N = 26) | *M* | 5.31 | 5.50 | 5.27 | 5.77 |
| | *SD* | 1.46 | 1.70 | 1.71 | 1.73 |
| Adaptive Top-down (N = 26) | *M* | 5.62 | 5.96 | 5.77 | 5.90 |
| | *SD* | 0.94 | 0.92 | 1.03 | 1.10 |
| Control (N = 26) | *M* | 5.27 | 5.27 | 5.15 | 5.38 |
| | *SD* | 1.64 | 1.69 | 1.71 | 1.65 |
| Overall (N = 130) | *M* | 5.38 | 5.58 | 5.46 | 5.70 |
| | *SD* | 1.40 | 1.48 | 1.51 | 1.54 |

## Exploratory Analyses

### *Feedback Experience Questionnaire*

Exploratory analyses were conducted on the Feedback Experience Questionnaire, which was given to participants after completing the final mission. This questionnaire asked participants about various aspects of the feedback given during the study. A repeated measures analysis was done on these feedback questions, and there was an interaction of questions and feedback condition, indicating that certain questions on the Feedback Experience Questionnaire were rated differently for different feedback groups. Therefore, separate analyses of individual questions were conducted, and post-hoc comparisons using the Tukey HSD test were conducted for significant $F$ values. Significant differences in responses were found for nine questions, and these are presented in Table 12. These differences show that people in the different feedback conditions were responding differently to the feedback. Overall, the feedback groups that included detailed feedback in some way (i.e., detailed, bottom-up, and top-down) rated these questions more favorably than the feedback groups that provided minimal information (i.e., general feedback group and the control group). The individuals in the detailed, bottom-up, and top-down groups tended to feel that the feedback more accurately reflected their performance, helped them to improve, and helped them generate learning strategies. These factors are all important to consider when designing feedback because feedback must not only provide adequate information when necessary, but an individual must be willing to accept the feedback and learn from it. The results of this exploratory analysis suggest that individuals believe that more descriptive feedback helps them learn better, and they are more willing to take and apply the information to help them perform better.

Table 12. Usability Questions and Significant Effects of Feedback Condition. Rating Scale is 1 (Strongly Disagree) to 6 (Strongly Agree).

| Question | Effect of Condition | Means (SD) | Tukey HSD post-hoc analysis |
|---|---|---|---|
| 1) The feedback I received was easy to understand | *F*(4, 109) = 3.491, *p* = .010, $\eta_p^2$ = .114, power = .848. | Detailed: 5.50 (0.76)<br>General: 4.35 (1.75)<br>Bottom-up: 5.12 (1.07)<br>Top-down: 5.27 (0.83)<br>Control: 5.31 (1.18) | Detailed (*p* = .006) and Top-down (*p* = .047) reported higher scores than General. |
| 2) The feedback correctly diagnosed errors. | *F*(4, 109) = 3.375, *p* = .012, $\eta_p^2$ = .110, power = .835. | Detailed: 5.08 (1.20)<br>General: 4.13 (1.60)<br>Bottom-up 5.04 (1.04)<br>Top-down: 4.88 (0.99)<br>Control: 3.85 (2.12) | Difference between the Detailed and control groups (*p* = .062) approached significance. |
| 3) Feedback helped me to improve my performance. | *F*(4, 109) = 7.495, *p* < .001, $\eta_p^2$ = .216, power = .996. | Detailed: 5.19 (0.94)<br>General: 4.00 (1.45)<br>Bottom-up: 5.00 (1.30)<br>Top-down: 4.65 (1.20)<br>Control: 3.00 (2.08) | Detailed (*p* = .021) reported higher scores than General.<br><br>Detailed (*p* < .001), Bottom-up (*p* <.001) and Top-down (*p* = .004) reported higher scores than the control group. |
| 4) Feedback helped focus my attention on learning strategies. | *F*(4, 108) = 6.764, *p* < .001, $\eta_p^2$ = .200, power = .992. | Detailed: 4.96 (0.96)<br>General: 4.09 (1.08)<br>Bottom-up 5.16 (1.14)<br>Top-down: 4.46 (1.24)<br>Control: 3.23 (1.96) | Bottom-up (*p* = .027) reported higher scores than General.<br><br>Detailed (*p* = .001), Bottom-up (*p* <.001) and Top-down (*p* = .033) reported higher scores than the control group. |
| 5) Feedback focused my attention towards goal performance level. | *F*(4, 107) = 4.476, *p* = .002, $\eta_p^2$ = .143, power = .931. | Detailed: 5.04 (0.96)<br>General: 4.35 (1.19)<br>Bottom-up: 5.04 (1.14)<br>Top-down: 4.54 (1.17)<br>Control: 3.42 (2.11) | Detailed (*p* = .003) and Bottom-up (*p* = .003) reported higher scores than the control group. |

| Question | Effect of Condition | Means (SD) | Tukey HSD post-hoc analysis |
|---|---|---|---|
| 6) Feedback could have been more useful. | $F(4, 108) = 3.555$, $p = .009$, $\eta_p^2 = .116$, power = .855. | Detailed: 3.08 (1.57)<br>General: 4.17 (1.44)<br>Bottom-up: 3.77 (1.42)<br>Top-down: 3.92 (1.32)<br>Control: 4.77 (1.36) | Detailed ($p = .006$) rated this question more favorably than the control group. |
| 9) I ignored the feedback. | $F(4, 108) = 3.233$, $p = .015$, $\eta_p^2 = .107$, power = .816. | Detailed: 1.38 (0.57)<br>General: 1.74 (0.69)<br>Bottom-up 1.38 (0.75)<br>Top-down: 1.92 (1.22)<br>Control: 2.31 (1.44) | Detailed ($p = .034$) and Bottom-up ($p = .034$) rated this question more favorably than the control group. |
| 10) Feedback provided me with effective strategies. | $F(4, 106) = 5.682$, $p < .001$, $\eta_p^2 = .177$, power = .976. | Detailed: 4.54 (1.10)<br>General: 3.43 (1.43)<br>Bottom-up: 4.50 (1.11)<br>Top-down: 4.20 (1.41)<br>Control: 2.85 (1.63) | Detailed ($p = .002$), Bottom-up ($p = .003$), and Top-down ($p = .026$) reported higher scores than the control group.<br><br>Detailed ($p = .037$) and Bottom-up ($p = .048$) reported higher scores than General. |
| 11) Feedback helped me generate my own strategies. | $F(4, 108) = 2.645$, $p = .037$, $\eta_p^2 = .089$, power = .722. | Detailed: 4.04 (1.46)<br>General: 3.50 (1.30)<br>Bottom-up: 4.58 (1.10)<br>Top-down: 4.00 (1.52)<br>Control: 3.23 (1.92) | Difference between Bottom-up and the control group ($p = .051$) approached significance. |

*Spatial Orientation*

Throughout the planned analyses, spatial orientation regularly correlated very highly with mission performance. Therefore, exploratory analyses were also conducted to determine the extent to which spatial orientation scores predicted variability in mission performance. A regression analysis was conducted using spatial orientation as a predictor of Mission 1 performance scores, and this analysis showed that spatial orientation predicts 14.5% of the variance in overall performance in Mission 1, $R^2 = .145$, $F(1, 128) = 21.709$, $p < .001$. A second regression analysis was conducted using spatial orientation as a predictor of Mission 2 performance, and this analysis demonstrated that spatial orientation predicts 14.7% of the variance in Mission 2 performance scores, $R^2 = .147$, $F(1, 128) = 22.108$, $p < .001$. A third regression analysis was conducted using spatial orientation as a predictor of Mission 3 performance, and spatial orientation accounts for 10.5% of the variance in Mission 3 performance scores, $R^2 = .105$, $F(1, 128) = 15.004$, $p < .001$. Finally, a fourth regression analysis was conducted using spatial orientation as a predictor of Mission 4 performance scores, and this analysis revealed that spatial orientation predicts 15.7% of the variance in Mission 4 performance, $R^2 = .157$, $F(1, 128) = 23.923$, $p < .001$.

These regression analyses indicate a significant relationship between spatial orientation scores and task performance. Spatial orientation is a substantial predictor of performance on all training missions. Taking this into consideration, it may be useful to pre-test individuals for spatial orientation and then design training according to their scores. However, more research is needed to determine exactly what kind of training would most benefit those individuals who score high on the spatial orientation measure as opposed to those who score very low.

# CHAPTER FIVE: DISCUSSION

Simulation-based training systems are already prevalent in training programs in various domains, and their use is likely to increase in the future (Chang, 2009). Following theoretically-based and empirically demonstrated guidelines for the design of adaptive feedback in these systems is very important. The present research was designed to provide empirical evidence for the efficacy of adaptive bottom-up feedback in SBT systems, based on several theoretical perspectives. First, the “2 sigma” problem was acknowledged, which is the consistent finding that one-to-one tutoring helps students achieve performance levels two standard deviations above levels in a traditional classroom environment (Bloom, 1984). A human tutor is able to adapt feedback to each individual student’s needs to help the student successfully learn the material. How can this idea of adaptive feedback be translated into SBT systems? Second, to explore this question, a feedback intervention using an adaptive bottom-up approach was designed, which aligned with the assumptions of the Cognitive Load Theory (CLT) and the Expertise Reversal Effect (ERE). Based on the CLT, adaptive feedback was designed to aid in schema acquisition while taking into account working memory limitations. Furthermore, based on the ERE feedback was given specific characteristics allowing it to adapt dynamically to each individual. These provided the adaptive bottom-up feedback framework that was used in the current study. In the context of this study, adaptive bottom-up feedback was defined as a feedback intervention where detailed feedback was given to each participant initially, followed by general feedback as an individual demonstrated competency in the task. Third, testable hypotheses were developed. In particular, it was hypothesized that formative feedback conditions would have learning benefits over outcome feedback alone. It was also hypothesized that the adaptive bottom-up feedback condition would be superior to the other formative feedback conditions in terms of performance levels, knowledge post-test scores, and cognitive load. These hypotheses were not fully supported. A summary of the results is provided, including a discussion of the interpretations and implications of these findings. Finally, limitations to the current study are discussed, followed by suggestions for future research.

## Summary of Results & Implications

The first hypothesis was that the four formative feedback groups (detailed, general, bottom-up, and top-down) would show learning benefits while the control group would not. The data show that there were significant performance improvements over missions for the detailed, bottom-up, and top-down conditions. These results indicate that significant learning occurred in these groups but not for the general feedback condition or the control group. Results also reveal that all formative feedback groups outperformed the control group on the transfer mission (Mission 4). This corroborates prior research showing that formative feedback is better than outcome feedback alone (Gilman, 1969; Astwood et al., 2008). A learning benefit for several feedback conditions is also seen for knowledge post-test scores. The detailed condition and both adaptive feedback groups scored higher than the control group on the knowledge post-test. Clearly, these approaches produce superior results when compared to the control group on knowledge post-test scores.

The second hypothesis was that the adaptive bottom-up feedback intervention would show benefits over the other three formative feedback groups in performance, retention, and cognitive load. Overall, results indicate that the detailed and bottom-up conditions are the best in terms of performance levels across the missions. Therefore, adaptive bottom-up feedback may not necessarily be the best approach if detailed feedback can be used to achieve the same results. However, the trend in data indicated that the bottom-up group performed better than the constant detailed group during training, although not significantly so. This suggests that bottom-up feedback may have other benefits in terms of performance and learning that were not specifically investigated in the current research effort.

Results also established that the bottom-up group performed consistently better than the general group on the training and transfer missions as well as the knowledge post-test. Moreover, the bottom-up condition performed significantly better than the top-down condition during the training missions, even though both adaptive groups performed at equal levels on the final transfer mission. Therefore, while the adaptive feedback groups ultimately achieved the same performance level, the bottom-up group got there faster. This has significant implications because the educational effectiveness of a training system can be measured by examining learning rate as a function of cost (Corbett, Koedinger, & Anderson, 1997). If a training system is effective, then it will allow students to reach higher performance levels in the same amount of time, or in less time, as other training methods. In this particular case, adaptive bottom-up or constant detailed feedback interventions would be better than implementing adaptive top-down feedback because shorter training times are needed to achieve the same level of performance. This also may affect long term retention of information, though this was not examined in the current study. If information is learned more quickly, there is more time to rehearse it; therefore, it may be retained for a longer period of time. Additionally, the differences between the two adaptive feedback groups may also explain why it took participants in the top-down group longer to perform at a comparable high level. With adaptive bottom-up feedback, detailed information helps an individual formulate ideas and concepts. With adaptive top-down feedback, an individual may form the wrong concept if not corrected in detail immediately. In terms of cognitive load, the results indicate that there are no significant differences between feedback groups for any of the missions; therefore, the cognitive load measure that was used may not have been sensitive enough to detect any differences.

While the predictions were not all supported, these results present empirical evidence for the Cognitive Load Theory, where more detailed feedback aids a novice in creating schemata and learning new information. Because no significant differences existed between the detailed and bottom-up conditions, we cannot conclusively state that evidence supporting the Expertise Reversal Effect (ERE) was found. According to this phenomenon, detailed feedback is beneficial for novices; however, as knowledge level increases, detailed feedback can cause performance decrements. If this effect had been evident, the bottom-up condition should have displayed significant benefits over all other conditions, including the detailed group. There are several possible explanations for this. Perhaps participants did not reach high enough levels of expertise in the time allotted for the experiment. It is possible that participants were still forming connections between bits of information, even as the experiment ended. With a longer training period and additional missions, the ERE may have been detected. Also, it is possible that the ERE was not detected because the experimental task was not cognitively demanding enough. In addition, the cognitive demand of processing detailed versus general feedback messages may not have been that different.

In conclusion, adaptive bottom-up feedback and constant detailed feedback can both be effectively applied to SBT exercises of short duration. In addition, these two feedback interventions provide a quicker and more efficient way to train a task, when compared to adaptive top-down feedback. While participants in the detailed, bottom-up, and top-down conditions attained the same levels of performance by the end of the training, the detailed and bottom-up groups attained that level of performance much more quickly. Furthermore, when comparing the detailed versus the bottom-up feedback conditions, the trend indicates that the bottom-up feedback provided additional benefits in terms of learning speed.

## Study Limitations

Several limitations of the current research should be addressed. The first limitation involves the mastery-based criteria used for the adaptive feedback conditions. In order to implement adaptive feedback in the simulation, it was necessary to determine the performance levels at which feedback specificity should be altered. No existing criteria for triggering this change in SBT was found in the literature, so steps were taken to ensure that the criteria chosen in this study were based on concrete principles. For the bottom-up condition, criteria were derived from pilot study data. The specificity of the feedback was determined by whether or not a participant met a certain criterion score (obtained from median scores from a pilot study). For the top-down condition, the criteria were based on the definition of the top-down approach to implementing feedback: Start a person with general feedback and continue giving them general feedback unless the person fails to show performance improvement (at which point, detailed feedback should be given). It is very possible that these were not the most appropriate criteria to use for the current research, yet no other research documents prescriptive ways to choose criteria for adaptive feedback.

The second limitation of this research involves the duration of training. The experiment consisted of several short training manuals followed by four missions that were each 10 minutes long. This may not have been adequate time for participants to learn the material and perform at high levels. Given more time and more missions, other significant differences between feedback conditions may appear. It may also be wise to examine longer training programs versus short experiments because participants may try to do better (or even worse) because it is a short experiment and has no real consequence in their lives (van Merriënboer & Sweller, 2005). In addition, the knowledge post-test was given immediately following the fourth mission. Different results may have been found if the post-test was given after a longer period of time (i.e., days, or weeks). Currently, we cannot conclusively identify the best feedback approach in terms of long term retention.

A third limitation involves the broad range of knowledge levels not specifically examined in this research. The current study defined a novice as an individual who knew a minimal amount of information about the task. An individual was assumed to have demonstrated competency in the experimental task when a certain performance criterion was reached. However, a novice does not simply jump from knowing nothing about a task to being competent in the task. According to Chi (2006), a novice must experience a wide range of proficiency levels before achieving complete mastery, and the current research did not take all of those levels into account. Hoffman (1998, as cited in Chi, 2006) identified a proficiency continuum with seven different levels that an individual can achieve (See Figure 8). First, an individual can be naïve, meaning that he or she is completely ignorant of a domain and his or her level of proficiency is extremely low. Then as a novice, an individual receives minimal exposure to the domain. At an initiate level, an individual begins introductory instruction. An individual achieves an apprentice level of proficiency when he or she is actively going through instruction. At a journeyman level, an individual can perform unsupervised, as he or she has become competent. When an individual reaches an expert proficiency level, he or she exhibits accurate and reliable performance. Finally, a master characterizes an expert who is highly proficient and able to effectively teach his or her skills or knowledge to others.

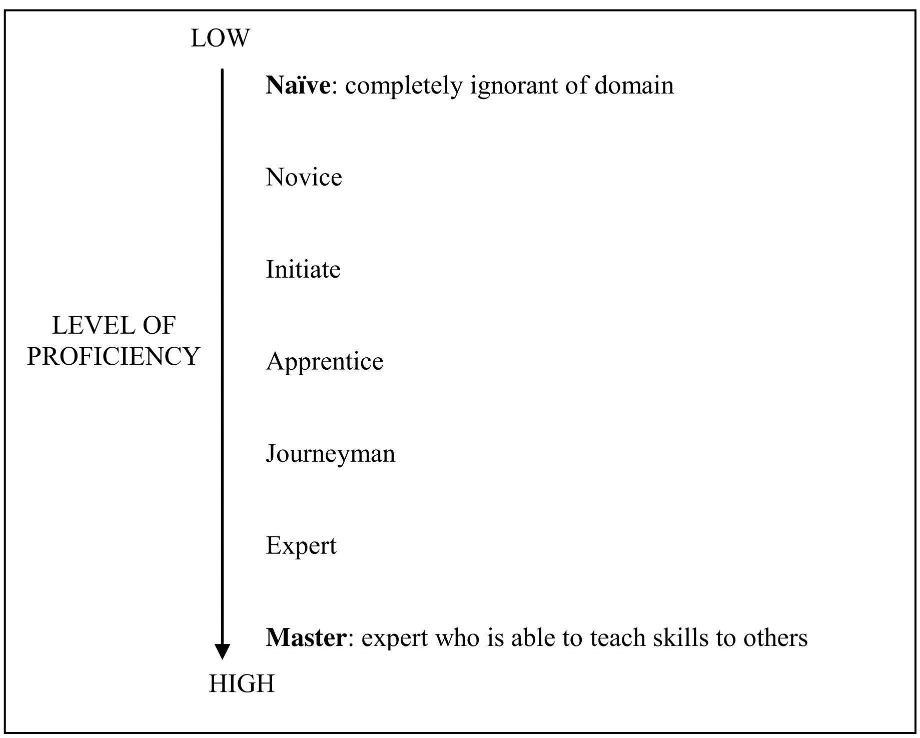



Figure 8. Levels of Proficiency, or Knowledge Level, on a Continuum (Adapted from Chi, 2006 and Hoffman, 1998).

Considering this range of proficiency levels may have been valuable in the current research in terms of determining how to fine tune different feedback specificities to multiple levels of knowledge. Different types of feedback may have been more appropriate for certain proficiency levels. In the current research, feedback was categorized as either general or detailed. However in reality, detailed and general feedback can include varying degrees of information. For example, detailed feedback may contain only a few details, or it may consist of a full summary of the problem and solution. Both are detailed, but each contains a different amount of detailed information. Detailed feedback can indicate what action an individual should take to rectify a mistake, or at a deeper level this feedback can also include the reasoning behind the action. It is possible that different levels of feedback specificity may correspond with the different levels of proficiency, described earlier. For instance, perhaps an apprentice would benefit more from less guidance than an individual in the initiate stage, but more guidance than an individual in the journeyman stage. Although the proficiency continuum and additional degrees of feedback specificity were not addressed in the present study, this represents an area of research that should be pursued in the future in terms of more personalized methods for adapting feedback to an individual's knowledge levels and proficiency.

The Cognitive Load Questionnaire may also be a drawback in this research. This measure was comprised of a single question about participants' perceived levels of cognitive load. No significant differences were found between feedback conditions, and one reason may be due to the lack of sensitivity of this measure. While other measures such as the NASA Task Load Index may be longer and more intensive to administer, they may have been more suitable for this study.

This deserves further investigation because it is also possible that the experimental task may not have been difficult enough to impose significant levels of workload for the participants.

Finally, the last limitation is related to the authenticity of the simulated task. While the search and rescue task was modeled from an exercise used in previous research (Oden, 2008) and contained some simple elements involved in real military training exercises, this was not a real military search and rescue task and may not have been challenging enough for participants. Therefore, the results have limited generalizability, although the results are expected to generalize to tasks of similar characteristics that require learning, integrating, and application of correct procedural information in various situations, where participants are left on their own to interpret computer-generated feedback. However, do the findings generalize to more cognitive or complex tasking? Just because adaptive bottom-up feedback works well in this domain does not necessarily mean that it will translate in the same way for all other domains. In addition, the results may not generalize beyond the age group of the participants, who were mostly college-age individuals.

## Future Research

Several researchers have noted that while we have made important advances in training research (cognition, training design, and training effectiveness), these practices still need to be incorporated into the design of new simulations (Salas & Cannon-Bowers, 2001). The research does no good if it is not applied, or if it is applied incorrectly. The current research focused on a theoretically-based implementation of adaptive feedback. The transitions in feedback specificity were based on on-going performance assessments that reflected an individual's knowledge level.

It would be interesting to directly compare this adaptive mastery-based feedback to a structured, time-based feedback method to see where differences may lie and if one method is truly better than the other. If adaptive mastery-based feedback has no true benefits over time-based feedback, then time-based feedback may be a better alternative because it is easier to implement (even though it is not truly individualized).

If adaptive mastery-based feedback continues to demonstrate learning benefits, researching ways to automate performance assessment may be beneficial. Mory (2004) noted that "Adaptive feedback information can easily be facilitated within a computer-based instruction environment where the computer can record and analyze the types of errors being made and give appropriate feedback based upon error types" (p. 758). This may be true for concrete domains (i.e. science and mathematics), but it is rarely found in SBT systems that involve scenario-based exercises and more abstract concepts. Most SBT systems are not programmed to automatically assess a wide variety of performance measures. Many times an instructor must be present to monitor performance and give feedback. While the technology to do this may exist, research is needed to determine how to automate performance measures in SBT systems, where errors in thinking may not be as easy to assess (Mangos & Johnston, 2009). This way, performance can be measured and feedback can be adapted to that performance automatically, creating much less workload for an instructor.

Future research should continue to examine the best ways to implement adaptive feedback in SBT systems. Specifically, more research is needed to determine if adaptive feedback has any significant benefits over constant detailed feedback in varying situations. In addition, research should investigate appropriate criterion levels to use in adaptive feedback instead of using arbitrary numbers, or numbers gathered from pilot studies. Adaptive feedback has the potential to be a great addition to SBT system architecture, if employed properly.

In addition to adapting feedback based on performance levels, future research should look at other alternative ways of adapting feedback to the individual. Individual characteristics may give certain people an advantage in SBT systems and may affect how different feedback interventions influence learning (aptitude-treatment interactions). Measuring for these characteristics during pre-training and subsequently tailoring training based on those characteristics may provide many training advantages. Pre-training could look at learner variables such as cognitive abilities, metacognitive skills, affective states (motivation, attention), personality, learning styles, etc. Then, based on these, instructional variables such as feedback type, timing, content sequencing, and rewards could be developed to enhance the training experience for that individual (Shute & Zapata-Rivera, 2007). For example, perhaps individuals can be pre-tested for aptitude. It is possible that high-ability students may benefit from less explicit feedback while low-ability student may benefit from more explicit and detailed feedback. Cognitive load could also be measured throughout training, and feedback could adapt to changes in load in an attempt to alleviate workload. Another alternative method is to give control of feedback to the students. In this way, they can determine when they need help most and then choose when to receive feedback. However, there is a caveat: it has been shown that this type of implementation can be abused by students who only want to complete the training rather than to learn the material (Shute, Woltz, & Regian, 1989). Giving feedback control to the student may not be a particularly wise decision, although more research is needed.

The current research takes a very theory-driven approach in that adaptive feedback was designed and implemented based on the Cognitive Load Theory. It would also be interesting to see if the findings hold true for human tutors as they instruct and give feedback to their students. If the findings do not hold true, it would be beneficial to examine the ways in which human tutors use different approaches and feedback strategies to enhance learning (e.g., how do they filter information, and how do they choose which points to emphasize?). By making such observations, we may gain insight on other effective strategies to include in the design of SBT systems.

Finally, the timing of feedback is another element of instruction that should be examined. In the current experiment, feedback was given immediately after each mission. This was in accordance with the concept of transfer appropriate processing, which emphasizes the need to keep the simulated task as close to the real-world task as possible. In the real world, an individual may not have access to immediate feedback, or coaching, or any feedback at all. However, some people have advocated coaching, or presenting feedback in real-time when people make mistakes. No consensus has been reached regarding the best time to provide feedback, so this is an area that deserves additional research.

## Conclusion

This study investigated the effects of different feedback interventions on performance, retention, and cognitive load in a simulated search and rescue task. In particular, the efficacy of adaptive bottom-up feedback was examined. While the predictions were not fully supported, adaptive bottom-up feedback proved to be a viable method of implementing feedback in the SBT system. Further research is needed to determine the extent of related benefits that this method provides.

The key contributions of this research lie in three main areas. First, adaptive feedback conditions using mastery-based criteria and a theoretical framework were developed and used in a SBT system. Previously, other studies did not implement a mastery-based performance criterion to determine presentation of adaptive feedback, and support for a theoretically based implementation of adaptive feedback was lacking. Second, this research indicated that feedback is an extremely important element in SBT systems, which aligns with prior research findings. Third, the results indicated that the adaptive bottom-up feedback is superior than the adaptive top-down group in terms of learning speed. Fourth, this research demonstrated that the adaptive bottom-up group produced performance comparable to the constant detailed (non-adaptive) condition, and both are beneficial ways to present feedback to individuals. In addition, the trend shows that the bottom-up feedback condition may have other advantages such as learning speed, meaning that this feedback intervention should be implemented if possible. If the benefits of one-to-one tutoring can be harnessed using appropriately implemented adaptive bottom-up feedback, simulation-based training systems may offer a very efficient training tool.

In conclusion, many SBT systems have the capability of automating scenario generation and adaptive feedback generation, which can reduce instructor workload and more closely imitate one-to-one tutoring (Mangos & Johnston, 2009). Nonetheless, appropriate ways to implement these aspects of individualized instruction need to be theoretically based and empirically documented. The pattern of results from this research looks like it is important to get detailed feedback in the beginning (which supports the Cognitive Load Theory). Although it appears crucial to get detailed feedback at the beginning, whether or not it switches to general feedback later during training may not matter as much. However, the research revealed that if an individual does not get detailed feedback in the beginning, his or her performance will not improve as quickly. The contributions of this research can serve as a guideline for the future development and implementation of adaptive feedback in SBT systems as well as other similar computer-based training systems.

# APPENDIX A: FEEDBACK TERMINOLOGY

## Feedback Terminology

| Feedback Term | Context of current research | Other related terms in the literature |
|---|---|---|
| Adaptive Bottom-up | Presenting detailed feedback initially, then transitioning to general feedback as competency is demonstrated | - Fading feedback (Goodman & Wood, 2009) |
| Adaptive Top-Down | Presenting general feedback initially, then transitioning to detailed feedback if performance does not continue to improve across missions. | - Reverse fading (Goodman & Wood, 2009) |
| Confirmatory Feedback | Verification of a correct response. Praise after desired response; increases and maintains performance and morale | - Merrill, Reiser, Merrill, & Landes (1995)<br>- Mory (2004) |
| Corrective Feedback | Error correction with varying degrees of detail (not necessarily punishment based). Can offer suggestions for improvement. | - Error feedback (Merrill, Reiser, Merrill, & Landes, 1995)<br>- Corrective feedback (Bangert-Drowns et al., 1991; Kulhavy & Stock, 1989) |
| Detailed Feedback | Step-by-step, very specific information; culminates in an explicit answer or solution for the individual | - Process feedback (Delgado, 2005; Hattie & Timperly, 2007; Astwood et al., 2008)<br>- Elaborate feedback (Narciss & Huth, 2004; Smits et al., 2008)<br>- Directive feedback (Sanders, 2005; Shute, 2008)<br>- High specificity (Shute, 2008; Davis et al., 2005; Goodman & Wood, 2009). |
| Formative Feedback | Feedback that includes corrective information that goes beyond the information included in outcome feedback. | - Shute (2008)<br>- Elaborated feedback (Narciss, 2008) |
| General Feedback | Conceptual and broad in nature; can include hints that are given to nudge a person in the right direction without explicitly giving the answer to the problem | - Global feedback (Smits et al., 2008; Davis et al., 2005)<br>- Conceptual feedback (Hays et al., 2009, Cagiltay, 2006)<br>- Facilitative feedback (Shute, 2008)<br>- Hints-based feedback (Shute, 2008) |
| Mastery-based/ Adaptive Feedback | Feedback changes dynamically in response to performance on a training task | - Adaptive automation: Parasuraman, Mouloua, & Molloy (1996)<br>- Adaptive instruction/feedback (Park & Lee, 2004; Shute & Zapata-Rivera, 2007) |
| Outcome Feedback | Feedback gives trainee an idea of how they are performing; often a performance score or correct/incorrect response; extent to which they performed well. | - Knowledge of results (KR; Mory, 2004)<br>- Knowledge of performance (KP; Narciss, 2008)<br>- Verification feedback (Shute, 2008)<br>- Right/wrong feedback (Bangert-Drowns et al., 1991) |
| Time-based Feedback | Different feedback content is administered according to a pre-determined sequence, usually based on time | - Scaffolding (Jones & Fleischman, 2001; Sharma & Hannafin, 2007)<br>- Fading (Goodman & Wood, 2009; Jones & Fleischman, 2001; Kester & Kirschner, 2009)<br>- Reverse fading (Goodman & Wood, 2009)<br>- Sequencing feedback (van Duyne et al., 2001; van Gog et al., 2008) |

# APPENDIX B: SEARCH AND RESCUE LEARNING OBJECTIVES

## Search and Rescue Learning Objectives

| Learning Objective | LO Rationale | LO Procedures |
|---|---|---|
| 1. Entering and exiting buildings | Multiple searchers might (accidentally) be searching the same area at the same time, and this presents opportunities for several types of mistakes. Primarily it becomes difficult to know if a building has already been searched, or if a building is currently being searched. | 1. Before entering a building, walk around the entire building to make sure it is not tagged. You do not want to begin searching a building that is currently being searched or has already been searched.<br>2. All buildings will be entered and exited through the same doorway.<br>3. Before entering a building you will tag an area to the left of the door with a spray tag (if there is a window next to the door, apply the tag to the left of the window).<br>4. After completely searching and exiting each building, you will tag the area to the right of the door you entered a spray tag. You will see the spray tag that you left upon entering the building move to the right side of the door when you do this, signifying that you have searched this building. |
| 2. Searching buildings | Following rules during a missing persons search is important. Making the wrong choice can have costly consequences, such as wasted time, rooms getting searched more than once, and/or some un-searched rooms. | 1. Use the building search order rule to decide the order that you will search buildings. Start at the right-most building on the map and continue searching the remaining buildings in a counter-clockwise direction.<br>2. Once you enter a building, you will search through the building using the Right Turn Rule. Every room within a building should be searched.<br>3. If a building has multiple floors or is divided into multiple sections, you need to inform HQ when a section/floor is clear. EXAMPLE: "section clear"<br>4. Make sure you search every building completely. Otherwise you run the risk of (1) not finding the individual you are searching for, (2) not finding target items. |
| 3. Headquarters (HQ) communications | During a search the primary goal is to find the missing person. HQ Communications are extremely important in achieving this goal. It is also important to be looking for target items that might be helpful. Target items are typically things that the person was last seen with. | 1. Locate/report ONLY the target items that are specified by HQ, and once an item has been found, text HQ to report which item has been found and the building number where it was found. EXAMPLE: "case in 99"<br>2. After you report the item, wait until you have instructions from HQ before you continue with your search.<br>3. After exiting and tagging a building, you will report via texting to HQ. It should include the building number and the status of the building. EXAMPLE: "99 cleared"<br>4. It is important to keep track of time in a search and rescue task. Send status reports to HQ when you are 200s and 400s into your mission. You should tell HQ what building you are in at that time or what building you are going to. EXAMPLE: "in 99" OR "going to 99"<br>5. Report the location of medics to HQ. EXAMPLE: medic in 99 |

# APPENDIX C: MISSION INSTRUCTIONS

# MISSION: ALPHA SECTOR

An Alzheimer's patient (see photo below) has gone missing. He was last seen carrying a case (see photo below). Any cases you see in the buildings should be considered "target" items for this search. Also, the terrorists may have placed bombs in the buildings to disperse the biological agent. These bombs are also considered "target" items for this search (see below).

| | | |
|---|---|---|
|  Missing Alzheimer's patient |  Case |  Bomb |

Reference the map on your right.

Your assigned area of responsibility for this search includes buildings: **33, 34, and 45**. Your start location is denoted on the map with the blue starburst.

Once the mission begins, a timer will automatically appear on the right side of the screen. You will have 600 seconds on the timer [10 minutes] to complete your search.

When you complete your search, text message HQ: **Mission complete**

Otherwise, HQ will send you a message when time is up.

**When you are ready to begin this mission:**

1. Send a text to HQ that says: **Ready**.
2. Wait for HQ to reply to your text with one that says, **"FO1: begin mission alpha!"**
3. Then, the timer will start and you may begin the mission.

# APPENDIX D. FEEDBACK MESSAGES

## Feedback messages for Learning Objective 1: Entering and exiting buildings.

| Rules of the LO | Detailed Feedback Message | General Feedback Message | Confirmatory (NO Errors on entire LO) |
|---|---|---|---|
| 1. Before entering a building, walk around the entire building to make sure it is not tagged. You do not want to begin searching a building that is currently being searched or has already been searched. | Before entering or tagging a building you should walk around the entire building to make sure it is not already tagged. | Remember to apply the procedures for entering and exiting buildings. | Good job applying the procedures for entering and exiting buildings! |
| 2. All buildings will be entered and exited through the same doorway. | Exit a building or a building section through the same door that you used to enter. | | |
| 3. Before entering a building you will tag an area to the left of the door with a spray tag (if there is a window next to the door, apply the tag to the left of the window). | Tag the area to the left of the door when you FIRST enter a building to begin your search. Do NOT tag each building section. | | |
| 4. After completely searching and exiting each building, you will tag the area to the right of the door you entered a spray tag. You will see the spray tag that you left upon entering the building move to the right side of the door when you do this, signifying that you have searched this building. | After completely searching and exiting a building you should tag the area to the right of the door that you tagged when you entered the building. Do NOT tag each building section. | | |

## Feedback messages for Learning Objective 2: Searching buildings.

| Rules of the LO | Detailed Feedback Message | General Feedback Message | Confirmatory (NO Errors on entire LO) |
|---|---|---|---|
| 1. Use the building search order rule to decide the order that you will search buildings. Start at the right-most building on the map and continue searching the remaining buildings in a counter-clockwise direction. | Start searching the right-most building on the map and search the remaining buildings in a counter-clockwise direction. | Remember to apply the correct procedures for searching buildings! | Good job applying the correct procedures for searching buildings! |
| 2. Once you enter a building, you will search through the building using the Right Turn Rule. Every room within a building should be searched. | Use the right turn rule to decide the order to search rooms (go right whenever there is a choice in direction). | | |
| 3. If a building has multiple floors or is divided into multiple sections, you need to inform HQ when a section/floor is clear.<br><br>EXAMPLE: “section clear” | If a building has multiple floors or multiple sections you should text HQ when a section or floor is clear. | | |
| 4. Make sure you search every building completely. Otherwise you run the risk of (1) not finding the individual you are searching for, (2) not finding target items. | Make sure you search every building completely. | | |

**Feedback messages for Learning Objective 3: Communicating with Headquarters.**

| Rules of the LO | Detailed Feedback Message | General Feedback Message | Confirmatory (NO Errors on entire LO) |
|---|---|---|---|
| 1. Locate/report ONLY the target items that are specified by HQ, and once an item has been found, text HQ to report which item has been found and the building number where it was found.<br><br>EXAMPLE: "case in 99" | Locate and report only the items specified by HQ. Include the item name and building number. | Remember to apply the correct procedures for HQ communications. | Good job on applying the correct procedures for HQ communications! |
| 2. After you report the item, wait until you have instructions from HQ before you continue with your search. | After you report an item you should wait until HQ responds before you continue with your search. | | |
| 3. After exiting and tagging a building, you will report via texting to HQ. It should include the building number and the status of the building.<br><br>EXAMPLE: "99 cleared" | Text HQ with the building number and status of the search (cleared) AFTER exiting and tagging a building. | | |
| 4. It is important to keep track of time in a search and rescue task. Send status reports to HQ when you are 200s and 400s into your mission. You should tell HQ what building you are in at that time or what building you are going to.<br><br>EXAMPLE: "in 99" OR "going to 99" | Send status reports to HQ at 200s and at 400s and include what building you are in or what building you are going to. | | |
| 5. Report the location of medics to HQ.<br><br>EXAMPLE: medic in 99 | Report the location of medics to HQ. Do not report the location of other civilians. | | |

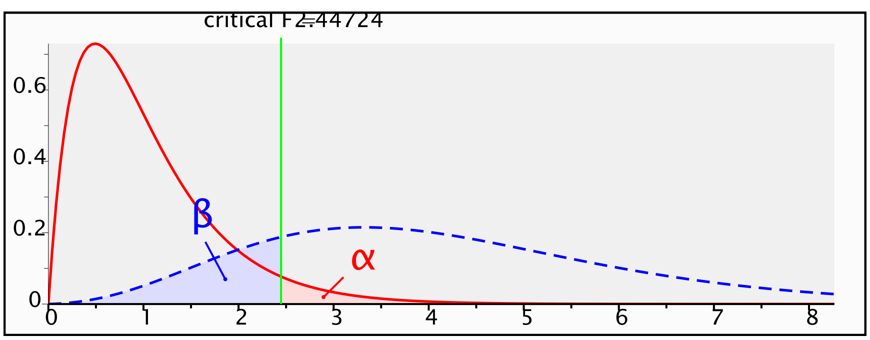
# APPENDIX E: POWER ANALYSIS

## Power Analysis for 5 x 4 mixed ANOVA using G*Power 3.

### Power Analysis: Between effects

**F tests -** ANOVA: Repeated measures, between factors

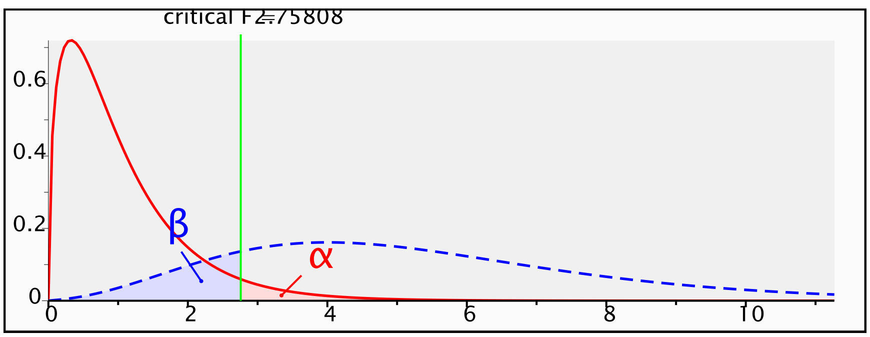
| | | | |
|---|---|---|---|
| **Analysis:** | A priori: Compute required sample size | | |
| **Input:** | Effect size f | = | 0.25 |
| | α err prob | = | 0.05 |
| | Power (1-β err prob) | = | 0.80 |
| | Number of groups | = | 5 |
| | Repetitions | = | 4 |
| | Corr among rep measures | = | 0.5 |
| **Output:** | Noncentrality parameter λ | = | 12.5000000 |
| | Critical F | = | 2.4472365 |
| | Numerator df | = | 4.0000000 |
| | Denominator df | = | 120 |
| | Total sample size | = | **125** |
| | Actual power | = | 0.8030360 |



## Power Analysis: Within effects

**F tests -** ANOVA: Repeated measures, within factors

| | | | |
|---|---|---|---|
| **Analysis:** | A priori: Compute required sample size | | |
| **Input:** | Effect size f | = | 0.25 |
| | α err prob | = | 0.05 |
| | Power (1-β err prob) | = | 0.80 |
| | Number of groups | = | 5 |
| | Repetitions | = | 4 |
| | Corr among rep measures | = | 0.5 |
| | Nonsphericity correction ε | = | 1 |
| **Output:** | Noncentrality parameter λ | = | 12.5000000 |
| | Critical F | = | 2.7580783 |
| | Numerator df | = | 3.0000000 |
| | Denominator df | = | 60.0000000 |
| | Total sample size | = | 25 |
| | Actual power | = | 0.8302870 |

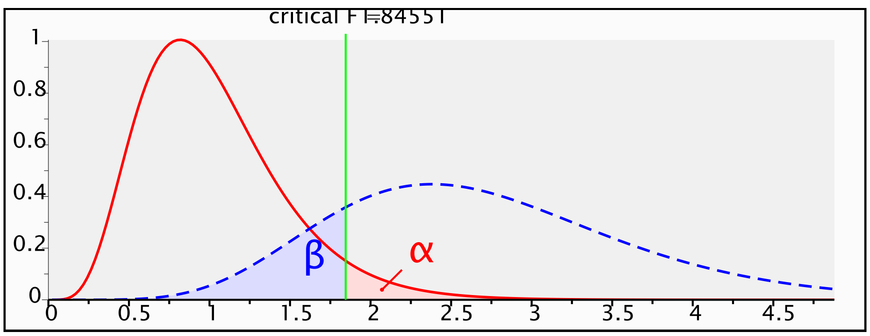


## Power Analysis: Between-within interaction effects

**F tests -** ANOVA: Repeated measures, within-between interaction

**Analysis:** A priori: Compute required sample size

| | | | |
|---|---|---|---|
| **Input:** | Effect size f | = | 0.25 |
| | α err prob | = | 0.05 |
| | Power (1-β err prob) | = | 0.80 |
| | Number of groups | = | 5 |
| | Repetitions | = | 4 |
| | Corr among rep measures | = | 0.5 |
| | Nonsphericity correction ε | = | 1 |
| **Output:** | Noncentrality parameter λ | = | 20.0000000 |
| | Critical F | = | 1.8455148 |
| | Numerator df | = | 12.0000000 |
| | Denominator df | = | 105 |
| | Total sample size | = | 40 |
| | Actual power | = | 0.8224234 |



# APPENDIX F: INFORMED CONSENT

**Informed Consent**

**Project title:** Simulation-Based Search and Rescue Training

*Please read this consent document carefully before you decide to participate in this study.*

**Age Requirement:** Only persons 18 – 64 years old may participate in this experiment.

**Purpose of the research study:** The purpose of this research is to examine search and rescue training in simulation-based training environments. This research will aid in the design of simulation environments in the future.

**What you will be asked to do in this study:** You will be asked to go through a training manual where you will learn how to use the simulation. You will also learn about the missions you will be going through and the guidelines for completing the missions. You will be a Forensics Officer and will control your movement through a simulated town as you complete different search and rescue missions at a computer. At the beginning of the study, you will be asked to complete several questionnaires. Then, following each mission, you will take a few additional short questionnaires about your experience.

**Time required:** We have allotted approximately 2 hours for your session; however, most people will complete the session in less than this time. You are free to terminate participation in this experiment at any time without bias. You will receive course credit or cash for the amount of time you participate or at least one hour.

**Benefits/Compensation:** As technology advances, different types of training and training elements are being used. In this study, we are trying to demonstrate that certain instructional design elements may be more beneficial to learning than others. The findings will serve as guidelines for the development of future training systems.

You will receive course credit for a research requirement, extra course credit (if offered by your instructor), or may elect to receive payment of $5 per half hour instead.

**Confidentiality:** Your identity will be kept confidential to the extent provided by law. Your information will be assigned a code number. The list connecting your name to this number will be kept in an electronic file. When the study is completed and the data have been analyzed, the list will be destroyed. Your name will not be used in any report. You will not be audio or video taped during the experiment.

**Voluntary participation:** Your participation in this study is voluntary. There is no penalty for not participating.

**Right to withdraw from the study:** You have the right to withdraw from the study at any time without consequences.

**Study contact for questions about the study or to report a problem:** If you have questions, concerns, complaints, or think the research has hurt you, talk to: Deb Billings, Graduate student, Applied Experimental and Human Factors Psychology Program, College of Sciences, 407-384-3592 or Dr. Richard Gilson, Faculty Supervisor, Department of Psychology at 407-823-2755. You can also send questions to ARI_RES@conus.army.mil or Deb.Billings@gmail.com and reference project name "Search and Rescue."

**IRB contact about your rights in the study or to report a complaint:** Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901. You may also talk to them for any of the following:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You want to get information or provide input about this research.

UCF University of Central Florida IRB
IRB NUMBER: SBE-09-06359
IRB APPROVAL DATE: 9/16/2009
IRB EXPIRATION DATE: 8/17/2010

**Project title:** <u>Simulation-Based Search and Rescue Training</u>

Having full capacity to consent, I do hereby volunteer to participate in research entitled Simulation-Based Search and Rescue Training under supervision of the U.S. Army Research Institute. I have been given an opportunity to read and keep a copy of this Agreement and to ask questions concerning this research. Any such questions have been answered to my full and complete satisfaction. I understand that I may at any time during this research revoke my consent and withdraw from the test without prejudice, and I will receive credit for the amount of time I participate or at least 30 minutes. I also understand that the rate of credit per hour (if applicable) will be determined by the instructor of the course for which the credit is to be applied. Alternatively, I may choose to receive $5 per 30 minutes in lieu of extra credit.

Please check how you would like to be compensated:

☐ I elect to receive course credit.

☐ I elect to receive $5/half hour for my participation.

Your signature below indicates your permission to take part in this research.

________________________________________
Name of participant (printed)

________________________________________ ____________________
Signature of participant Date

☐ I am at least 18 years of age or older.

**DO NOT SIGN THIS FORM AFTER THE IRB EXPIRATION DATE BELOW**

________________________________________ ____________________
Signature of person obtaining consent Date

________________________________________
Printed name of person obtaining consent

UCF University of Central Florida IRB
IRB NUMBER: SBE-09-06359
IRB APPROVAL DATE: 9/16/2009
IRB EXPIRATION DATE: 6/17/2010

# APPENDIX G: DEMOGRAPHICS QUESTIONNAIRE

## Demographics Questionnaire

1a. Year of birth: ______________ 1b. Gender: Male ______ Female ______

2. Have you graduated from high school? Yes______ No______

3. Which hand do you write with? Right______ Left______

4. Is your vision in each eye correctable to 20/20? Yes______ No______

5. To your knowledge, are you color blind? Yes______ No______

6. Do you own or have access to a computer? Yes______ No______

7. If yes, how often do you use a computer?
Daily_____ Several times a week ______ Occasionally_______ Never_______

8. Estimate how many hours per week you use a computer (circle one).

| | | | | |
|---|---|---|---|---|
| 0-9 hours | 10-19 hours | 20-29 hours | 30-39 hours | 40+ hours |

9. How do you rate your computer skills?
Novice/Beginner______ Intermediate______ Expert_______

10. Do you use the Internet? Yes______ No______

11. Do you own or use a video game system? Yes______ No______

12. How would you rate your video game skills?
Novice/Beginner______ Intermediate______ Expert_______

13. What is your level of confidence with video games in general?

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| Low | | Average | | High |

14. How many hours per week do you <u>currently</u> play video games?

| | | | | |
|---|---|---|---|---|
| 0-9 hours | 10-19 hours | 20-29 hours | 30-39 hours | 40+ hours |

15. How often do you play first person shooter games (e.g., Half-Life, Unreal)

| | | | | |
|---|---|---|---|---|
| Never | Rarely | Monthly | Weekly | Daily |

# APPENDIX H: KNOWLEDGE PRE-TEST AND POST-TEST

**Knowledge pre-test**

*Circle "T" for each item that is true and "F" for each item that is false.*

1. T / F On a search and rescue mission, all suspicious items should be reported immediately to Headquarters (HQ).

2. T / F When searching a building, begin with the top floor and work your way down.

3. T / F You do not need to fully enter a room to clear it.

4. T / F During a search and rescue mission, you should NOT report both the item and the item's location to HQ—this is redundant information because HQ already knows where you are.

5. T / F All rooms on a particular floor or section need to be searched before going to another floor or section.

6. T / F As long as you successfully clear a building, it does not matter which doors you enter and exit from.

7. T / F You should walk around the entire exterior of a building before you go inside to search it.

8. T / F When searching rooms, you should turn right whenever there is a choice in direction.

9. T / F Use the building size rule when deciding which building to search first. Smaller buildings should be searched first.

10. T / F You should keep track of your time in search and rescue missions.

**Knowledge post-test**

*Circle "T" for each item that is true and "F" for each item that is false.*

1. T / F If you see a building that is tagged to the left of the doorway, this means that someone is currently searching that building.
2. T / F You should report to HQ any time you exit a building, regardless of whether or not you have completely cleared it.
3. T / F You should give location updates to HQ at 100s and 400s.
4. T / F Search the building that is the closest to you first.
5. T / F You need to search each room twice to make sure it is completely clear.
6. T / F You need to stop your search and wait for orders from HQ when you find target items.
7. T / F Target items and other suspicious looking objects need to be reported immediately to HQ.
8. T / F If a building has several different sections, you should tag each section and report to HQ when you clear it.
9. T / F You need to report the locations of any medics or civilians you see during your search.
10. T / F You need to report every action that you take to HQ so that they will remain aware of how the search is progressing.

# APPENDIX I: FEEDBACK EXPERIENCE QUESTIONNAIRE

**FEEDBACK Experience Questionnaire**

**During the training phase, did you receive feedback from the experimenter after each trial?**

YES NO

***If YES, please continue. If NO, skip to question 12.***

**Please think about the feedback you received during the training and indicate on the scale from 1-6 your level of agreement or disagreement with the following statements.**

| | Strongly Disagree | | | | | Strongly Agree |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** |
| 1. The feedback I received was easy to understand. | | | | | | |
| 2. I believe that the feedback I received correctly diagnosed the errors I was making. | | | | | | |
| 3. I believe that the feedback I received helped me to improve my performance on the subsequent trials. | | | | | | |
| 4. I believe that the feedback I received focused my attention on learning strategies to perform this task better. | | | | | | |
| 5. I believe that the feedback I received focused my attention toward the performance level I should obtain. | | | | | | |
| 6. I believe that the feedback I received could have been more useful. | | | | | | |
| 7. It seemed like I received the same feedback over and over. | | | | | | |
| 8. I believe that the feedback I received did not accurately reflect my performance. | | | | | | |
| 9. I ignored and made no attempt to use the feedback I had received. | | | | | | |
| 10. I believe that the feedback I received provided me with effective strategies to help me perform better. | | | | | | |
| 11. I believe that the feedback I received helped me generate my own strategies to help me perform better. | | | | | | |

***Skip to Question 16***

**ONLY ANSWER THE FOLLOWING QUESTIONS IF YOU DID NOT RECEIVE FEEDBACK.**

**Please indicate on the scale from 1-6 your level of agreement or disagreement with the following statements.**

| | Strongly Disagree | | | | | Strongly Agree |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** |
| 12. I believe that feedback would have helped me improve my performance. | | | | | | |
| 13. I would have liked to have received feedback on my performance. | | | | | | |
| 14. I believe that having feedback would have motivated me more. | | | | | | |
| 15. I believe that having feedback would have increased my confidence more. | | | | | | |

**16. I have the following additional comments I would like to make concerning the feedback I was just provided with during this experiment.**

______________________________________________________________________________

______________________________________________________________________________

______________________________________________________________________________

______________________________________________________________________________

# APPENDIX J: MANIPULATION CHECK PROCEDURE AND DETAILS

## Manipulation Check Procedure and Details

Analyses were performed using SPSS 14.0 and Statistica 7 for Windows. An alpha level of .05 was used for all analyses, unless otherwise noted. Before any analyses were performed, the data were examined for any issues that could potentially affect the results of the statistical analyses. First, a manipulation check was performed by examining data for each learning objective (LO) for each participant in the two adaptive feedback groups to make sure that all participants actually received adaptive feedback for the three learning objectives (i.e., feedback content switched between detailed and general at some point across the missions).

The adaptive bottom-up data were checked first (N = 26). Figure 1 shows frequencies for learning objective 1. Four participants did not receive any change in feedback on this LO, and four others only experienced transitions between detailed and positive feedback (i.e., they scored 100%). Figure 2 shows frequencies for learning objective 2 for the bottom-up group. One participant never experienced adaptive feedback, and four others only experienced transitions between detailed and positive feedback. Figure 3 shows frequencies for learning objective 3, and eight of the participants did not receive a change in feedback over missions.

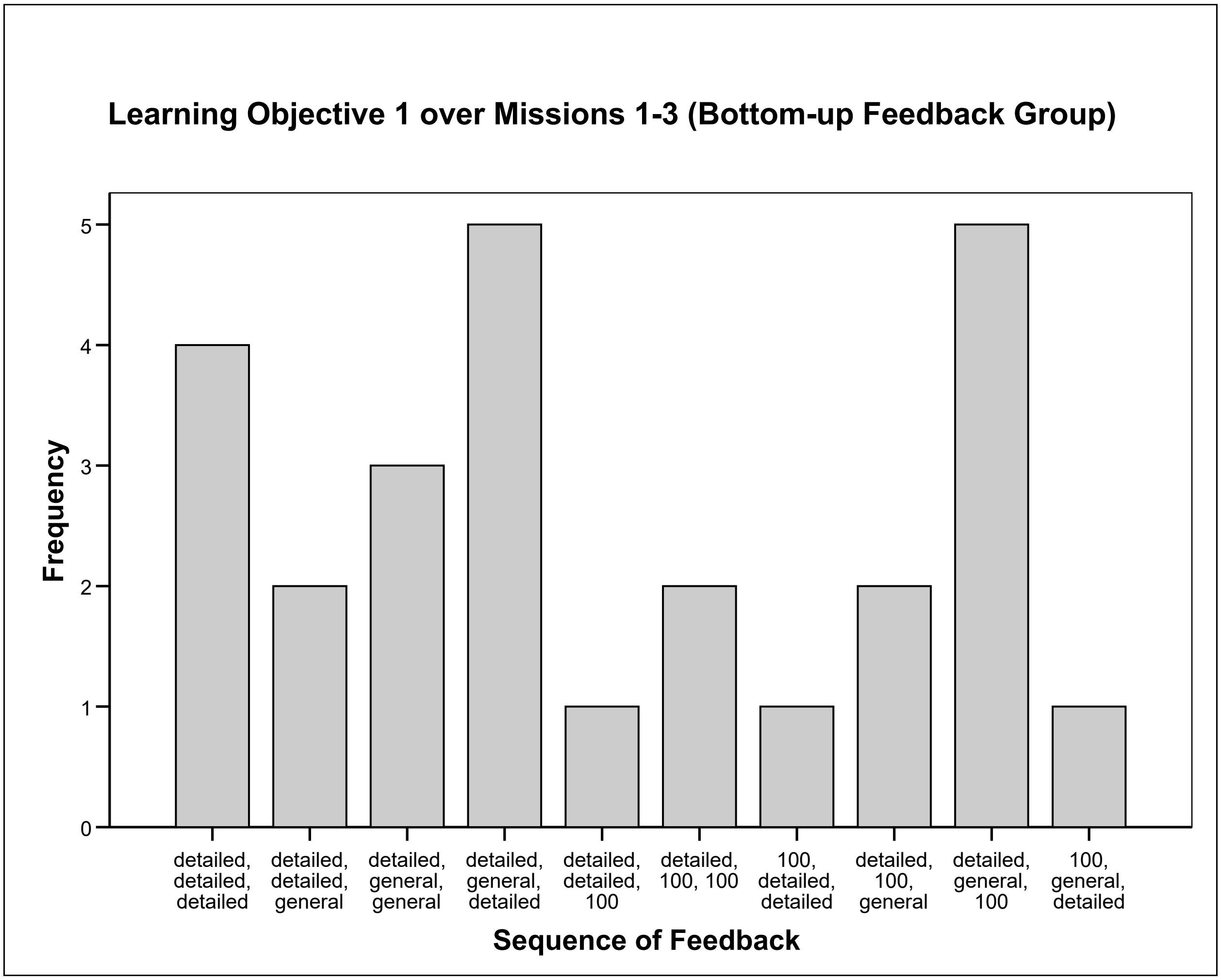



Figure 1. Bottom-up group: Feedback received for learning objective 1 across Mission 1, 2, and 3. Labels on the x-axis represent detailed feedback ("detailed"), general feedback ("general"), or positive feedback ("100"; i.e., score of 100% was obtained).

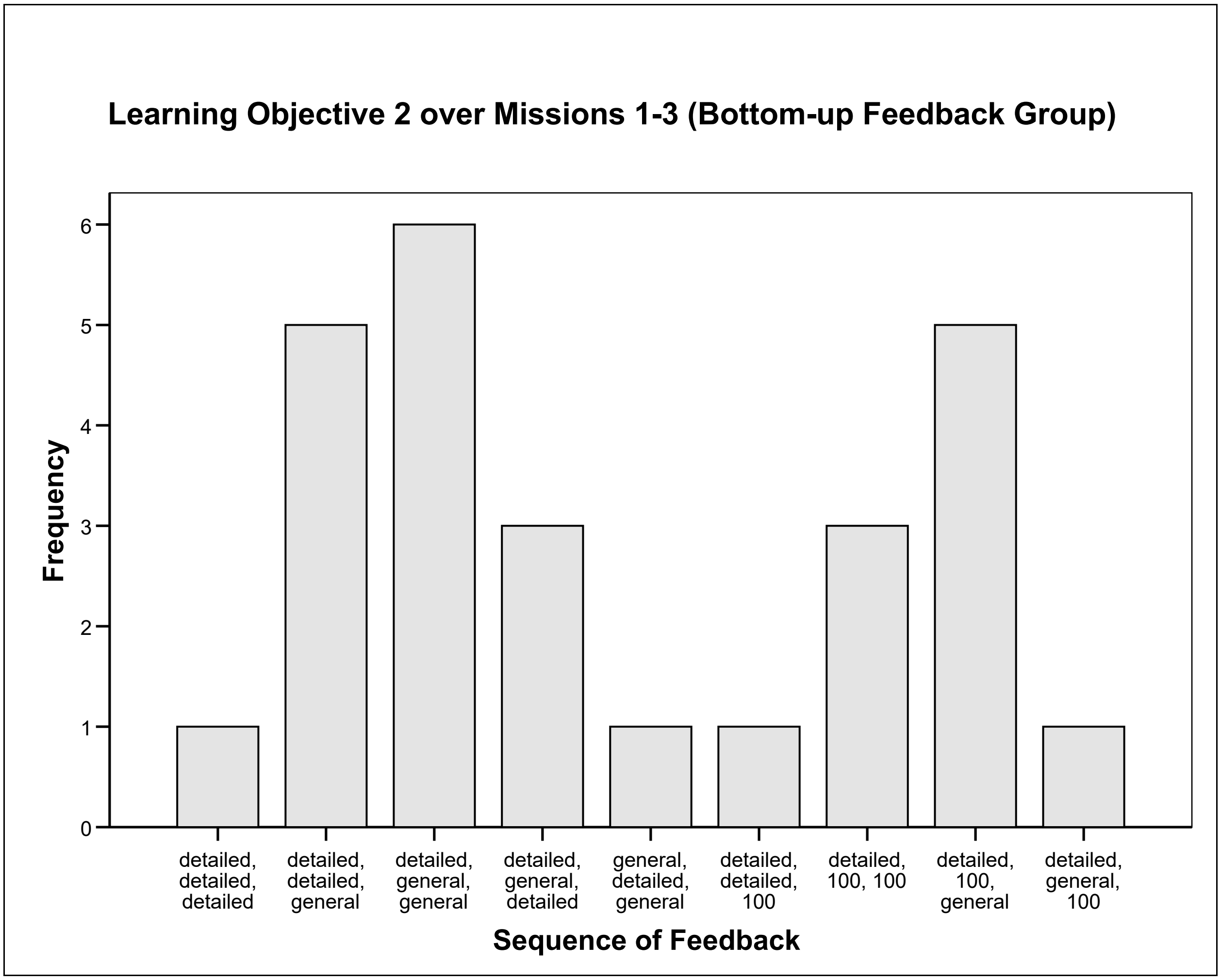



Figure 2. Bottom-up group: Feedback received for learning objective 2 across Mission 1, 2, and 3. Labels on the x-axis represent detailed feedback ("detailed"), general feedback ("general"), or positive feedback ("100"; i.e., score of 100% was obtained).

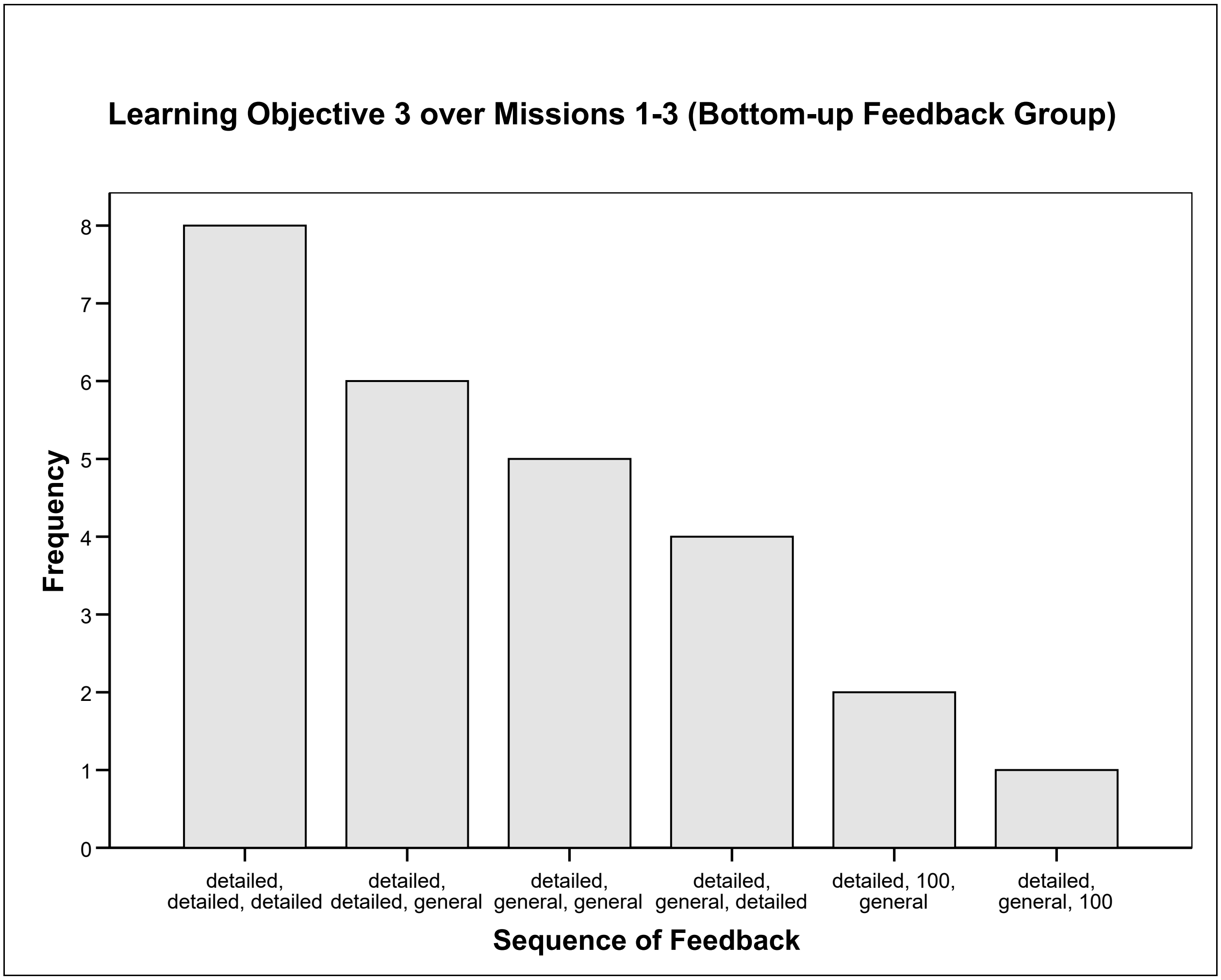



Figure 3. Bottom-up group: Feedback received for learning objective 3 across Mission 1, 2, and 3. Labels on the x-axis represent detailed feedback ("detailed"), general feedback ("general"), or positive feedback ("100"; i.e., score of 100% was obtained).

Next, data from the adaptive top-down condition were examined (N = 26). Figure 4 shows frequencies for learning objective 1, and one participant in this group did not receive any change in feedback. Two participants only experienced transitions from positive feedback to detailed feedback. Figure 5 shows frequencies for learning objective 2 for the top-down group. Three participants never received the change, and two others only experienced transitions between detailed and positive feedback, or between general and positive feedback. Figure 6 shows frequencies for learning objective 3, and one of the participants did not receive adaptive feedback. One other participant only experienced a change between positive and detailed feedback.

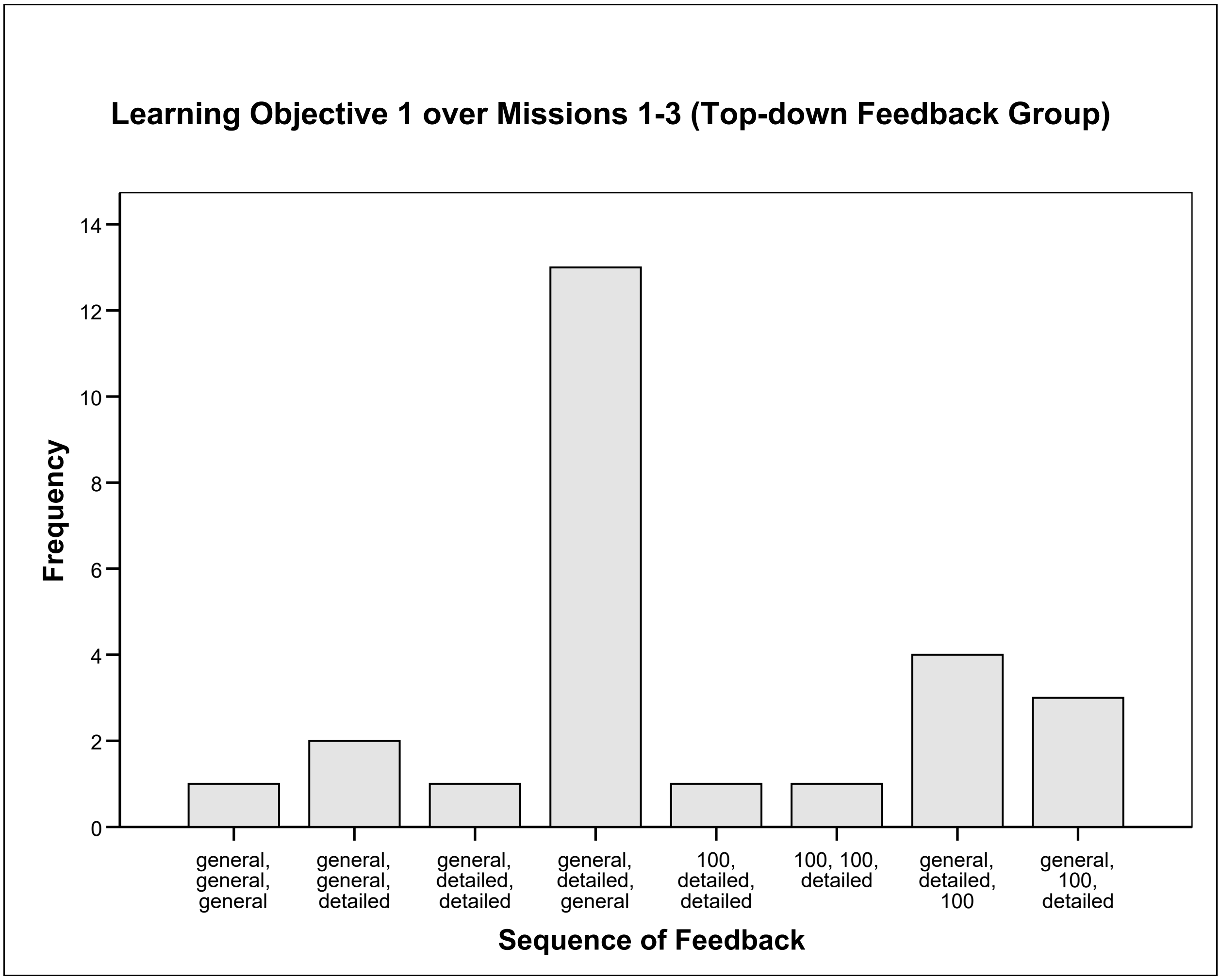



Figure 4. Top-down group: Feedback received for learning objective 1 across Mission 1, 2, and 3. Labels on the x-axis represent detailed feedback ("detailed"), general feedback ("general"), or positive feedback ("100"; i.e., score of 100% was obtained).

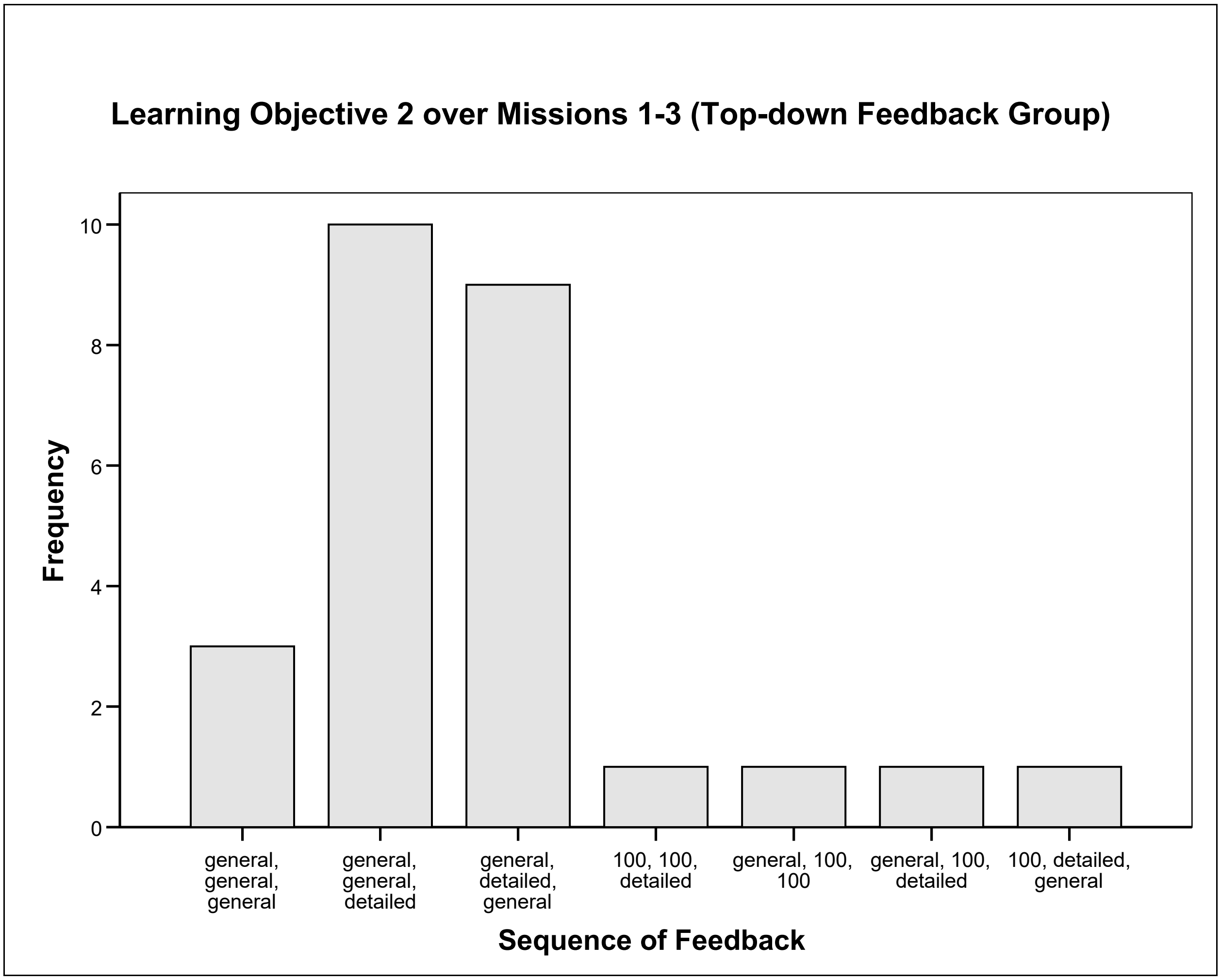



Figure 5. Top-down group: Feedback received for learning objective 2 across Mission 1, 2, and 3. Labels on the x-axis represent detailed feedback ("detailed"), general feedback ("general"), or positive feedback ("100"; i.e., score of 100% was obtained).

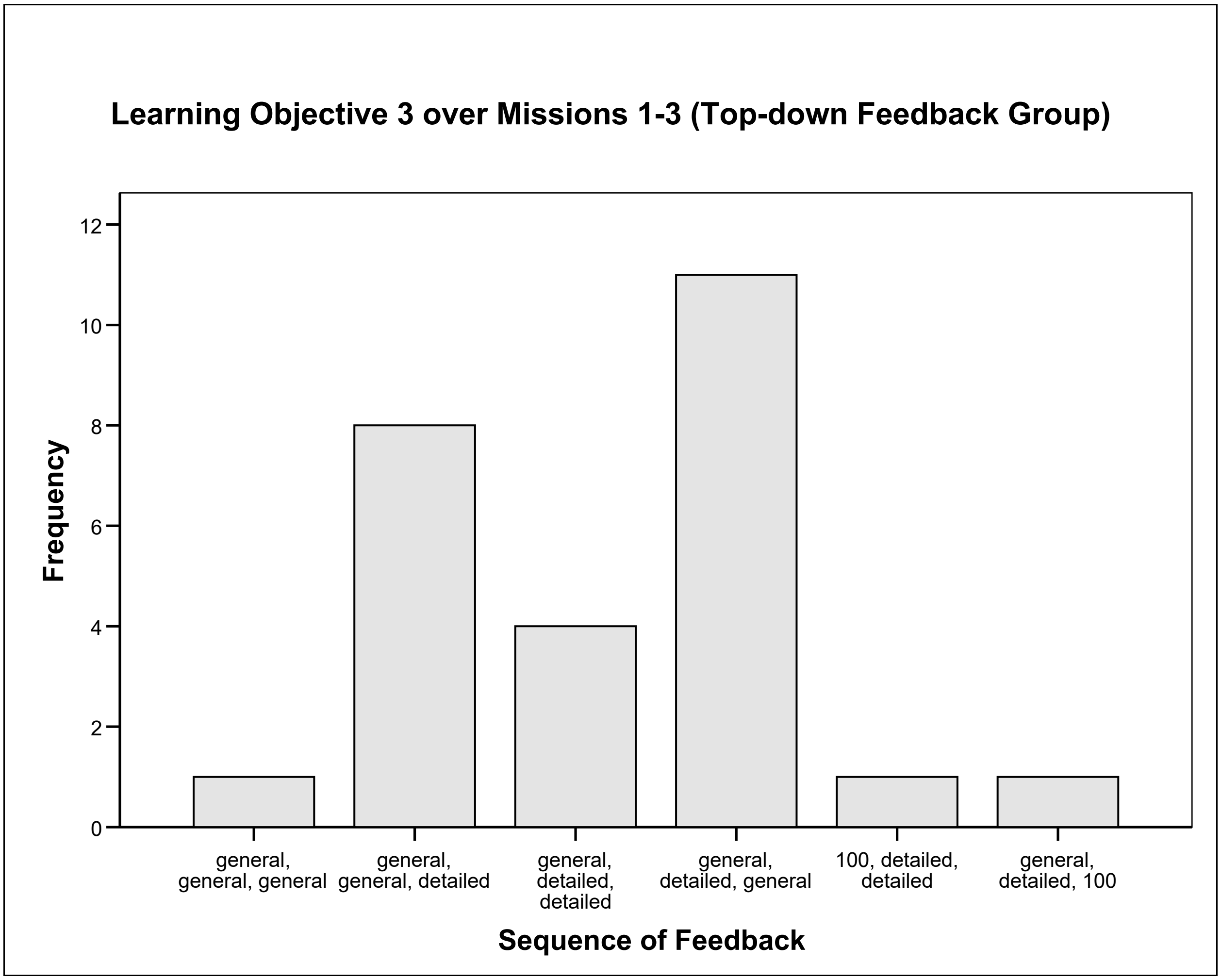



Figure 6. Top-down group: Feedback received for learning objective 3 across Mission 1, 2, and 3. Labels on the x-axis represent detailed feedback ("detailed"), general feedback ("general"), or positive feedback ("100"; i.e., score of 100% was obtained).

Although these findings indicate that every participant did not experience a change in feedback for all three learning objectives, it is still possible that each individual experienced adaptive feedback at some point during the training session as a whole, with changes occurring for some learning objectives but perhaps not for others. Therefore, the entire training session was examined. Findings confirmed that every participant in the bottom-up and top-down feedback conditions experienced the manipulation during training, although not every participant experienced it for all three learning objectives.

The data were also examined to determine the uniqueness of each individual's feedback experience in the two adaptive feedback conditions during the course of training. Feedback sequences over all missions and learning objectives were inspected. See Table 7 for feedback sequences for the bottom-up condition and Table 8 for feedback sequences for the top-down condition. For the bottom-up feedback group, 24 participants received unique feedback sequences, and two participants received identical sequences over the course of training. For the top-down feedback group, 17 participants received unique feedback experiences during training. However, several groups of individuals received identical feedback sequences (five participants, two participants, and two other participants). Although some participants experienced the same sequence of adaptive feedback during training, each individual's experience was based on his or her own performance during training. Hence, similar performance generated similar feedback sequences, which were personalized for each particular person.

Table 7. Feedback Sequences for the Adaptive Bottom-up Feedback Condition (N = 26).
d = detailed feedback, g = general feedback, 100 = positive feedback message [scored 100% on learning objective]

| | **Feedback Received** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Unique sequences | Learning Objective 1 | | | Learning Objective 2 | | | Learning Objective 3 | | |
| | **Mission 1** | **Mission 2** | **Mission 3** | **Mission 1** | **Mission 2** | **Mission 3** | **Mission 1** | **Mission 2** | **Mission 3** |
| (N = 1) | | | | | | | | | |
| 1 | d | g | d | d | 100 | g | d | d | d |
| 2 | d | d | g | d | g | d | d | d | d |
| 3 | 100 | d | d | d | d | g | d | g | d |
| 4 | d | 100 | g | d | 100 | g | d | d | d |
| 5 | d | 100 | 100 | d | g | g | d | g | 100 |
| 6 | d | g | g | d | 100 | 100 | d | d | g |
| 7 | d | g | g | d | g | d | d | 100 | g |
| 8 | d | d | 100 | d | g | g | d | d | d |
| 9 | d | d | d | g | d | g | d | d | g |
| 10 | d | g | 100 | d | 100 | 100 | d | 100 | 3 |
| 11 | d | d | d | d | d | d | d | d | g |
| 12 | d | d | g | d | d | g | d | d | d |
| 13 | d | d | d | d | d | g | d | d | d |
| 14 | d | g | g | d | g | 100 | d | d | d |
| 15 | d | g | d | d | g | d | d | g | g |
| 16 | d | g | 100 | d | d | 100 | d | g | g |
| 17 | d | g | 100 | d | d | g | d | d | d |
| 18 | d | d | d | d | 100 | g | d | d | g |
| 19 | d | g | d | d | g | g | d | d | g |
| 20 | d | 100 | 100 | d | 100 | g | d | d | g |
| 21 | 100 | g | d | d | 100 | 100 | d | g | d |
| 22 | d | g | 100 | d | g | g | d | g | g |
| 23 | d | g | 100 | d | 100 | g | d | g | d |
| 24 | d | 100 | g | d | d | g | d | g | d |
| (N = 2) | | | | | | | | | |
| 25 | d | g | d | d | g | g | d | g | g |

Table 8. Feedback Sequences for the Adaptive Top-down Feedback Condition (N = 26).
d = detailed feedback, g = general feedback, 100 = positive feedback message [scored 100% on learning objective]

| | **Feedback Received** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Unique sequences | Learning Objective 1 | | | Learning Objective 2 | | | Learning Objective 3 | | |
| | **Mission 1** | **Mission 2** | **Mission 3** | **Mission 1** | **Mission 2** | **Mission 3** | **Mission 1** | **Mission 2** | **Mission 3** |
| (N = 1) | | | | | | | | | |
| 1 | g | g | d | g | d | g | g | d | g |
| 2 | 100 | 100 | d | g | g | d | g | d | d |
| 3 | g | d | g | g | d | g | g | g | d |
| 4 | g | g | g | g | d | g | g | d | g |
| 5 | g | g | d | g | g | g | g | g | g |
| 6 | g | d | 100 | g | d | g | g | g | d |
| 7 | g | d | g | 100 | 100 | d | g | d | 100 |
| 8 | g | d | 100 | g | g | g | g | d | g |
| 9 | 100 | d | d | g | g | g | g | d | g |
| 10 | g | 100 | d | g | d | g | g | g | d |
| 11 | g | d | 100 | g | 100 | 100 | g | d | g |
| 12 | g | 100 | d | g | g | d | g | d | g |
| 13 | g | d | 100 | g | g | d | g | d | g |
| 14 | g | d | g | 100 | d | g | g | d | g |
| 15 | g | d | d | g | g | d | g | d | d |
| 16 | g | 100 | d | g | 100 | d | 100 | d | d |
| 17 | g | d | g | g | g | d | g | d | g |
| (N = 2) | | | | | | | | | |
| 18 | g | d | g | g | d | g | g | d | g |
| (N = 2) | | | | | | | | | |
| 19 | g | d | g | g | d | g | g | d | d |
| (N = 5) | | | | | | | | | |
| 20 | g | d | g | g | d | g | g | g | d |

## APPENDIX K: IRB APPROVAL LETTERS

## IRB Approval Letters

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

### Approval of Human Research

From: **UCF Institutional Review Board #1**
**FWA00000351, IRB00001138**

To: **Deborah Billings**

Date: **September 16, 2009**

Dear Researcher:

On 9/16/2009, the IRB approved the requested modifications, including revised Informed Consent document, until 08/17/2010 inclusive:

| | |
|---|---|
| Type of Review: | IRB Addendum and Modification Request Form |
| Project Title: | Adaptive feedback in simulation-based training. |
| Investigator: | Deborah Billings |
| IRB Number: | SBE-09-06359 |
| Funding Agency: | Consortium Research Fellows Program |
| Grant Title: | |
| Research ID: | N/A |

The Continuing Review Progress Report must be submitted 2 – 4 weeks prior to the expiration date for studies that were previously expedited, and 8 weeks prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 08/17/2010, approval of this research expires on that date.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a signed copy of the consent form(s).

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Joseph Bielitzki, DVM., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 09/16/2009 04:26:28 PM EDT

Joanne Muratori

IRB Coordinator

Page 1 of 1

University of Central Florida

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

# Approval of Human Research

From: **UCF Institutional Review Board #1**
**FWA00000351, IRB00001138**

To: **Deborah Billings**

Date: **October 26, 2009**

Dear Researcher:

On10/26/2009, the IRB approved the following modification for human participant research until 08/17/2010 inclusive:

| | |
|---|---|
| Type of Review: | IRB Addendum and Modification Request Form |
| Modification Type: | Replacing one study questionnaire with another |
| Project Title: | Adaptive feedback in simulation-based training. |
| Investigator: | Deborah Billings |
| IRB Number: | SBE-09-06359 |
| Funding Agency: | Consortium Research Fellows Program |
| Grant Title: | |
| Research ID: | N/A |

The Continuing Review Progress Report must be submitted 2 – 4 weeks prior to the expiration date for studies that were previously expedited, and 8 weeks prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 08/17/2010, approval of this research expires on that date.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Joseph Bielitzki, DVM., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 10/26/2009 10:41:32 AM EST

Joanne Muratori

IRB Coordinator

Page 1 of 1

## REFERENCES


Anania, J. (1983). The effects of quality of instruction on the cognitive and affective learning of students. Doctoral dissertation, The University of Chicago, Illinois. Retrieved October 2, 2009, from Dissertations & Theses: Full Text (Publication No. AAT T-28171).

Astwood, R. S., Van Buskirk, W. L., Cornejo, J. M., & Dalton, J. (2008). The impact of different feedback types on decision-making in simulation based training environments. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting,* 2062-2066.

Ayres, P., & van Gog, T. (2009). State of the art research into cognitive load theory. *Computers in Human Behavior, 25*, 253-257.

Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 4*, 417-423.

Baddeley, A.D., & Hitch, G. (1974). *Working memory.* In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory, Volume 8* (pp. 47-89). New York, NY: Academic Press.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.

Bannert, M. (2002). Managing cognitive load—recent trends in cognitive load theory. *Learning and Instruction, 12*, 139-146.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4-16.

Bolton, A. E. (2006). Immediate versus delayed feedback in simulation-based training: Matching feedback delivery timing to the cognitive demands of the training exercise. Doctoral dissertation, University of Central Florida. Retrieved July 20, 2009, from Dissertations & Theses (Publication No. AAT 3233644).

Borthick, A. F., Jones, D. R., & Wakai, S. (2003). Designing learning experiences within learners' zones of proximal development (ZPDs): Enabling collaborative learning on-site and online. *Journal of Information Systems, 17*(1), 107-134.

Bull, K. S., Shuler, P., Overton, R., Kimball, S., Boykin, C., Griffin, J. (1999). Processes for developing scaffolding in a computer mediated learning environment. In: *Rural Special Education for the New Millennium*. Conference Proceedings of the American Council on Rural Special Education (ACRES), Albuquerque, New Mexico, March 25-27.

Burke, A. J. (1984). Students' potential for learning contrasted under tutorial and group approaches to instruction. Doctoral dissertation, The University of Chicago, Illinois. Retrieved October 2, 2009, from Dissertations & Theses: Full Text.(Publication No. AAT T-28810).

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42,* 249-268.

Cagiltay, K. (2006). Scaffolding strategies in electronic performance support systems: Types and challenges. *Innovations in Education and Teaching, 43*(1), 93-103.

Camp, G., Paas, F., Rikers, R., & van Merriënboer, J. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior, 17*, 575-595.

Chang, H. (2009). Simulators always valuable in military training. Retrieved on January 20, 2010 from http://www.army.mil/-news/2009/04/13/19599-simulators-always-valuable-in-military-training/index.html

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Feltovich (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 21-30). Cambridge: Cambridge University Press.

Clark, R. C., & Mayer, R. E. (2008). Learning by viewing versus learning by doing: Evidence-based guidelines for principled learning environments. *Performance Improvement, 47*(9), 5-13.

Clark, R. C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. San Francisco, CA: Pfieffer, A Wiley Imprint.

Corbett, A. (2001). Cognitive computer tutors: Solving the two-sigma problem. In M. Bauer, P.J. Gmytrasiewicz, & J. Vassileva (Eds.), *UM 2001, LNAI 2109* (pp. 137-147). Berlin: Springer-Verlag.

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In *Proceedings of ACM CHI 2001 Conference on Human Factors in Computing Systems* (pp. 245-252). New York: Association for Computing Machinery Press.

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction, second completely revised edition* (pp. 849-874). The Netherlands: Elsevier Science B.V.

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. *Computers in Human Behavior, 20*, 225-241.

Davis, W. D., Carson, C. M., Ammeter, A. P., & Treadway, D. C. (2005). The interactive effects of goal orientation and feedback specificity on task performance. *HumanPerformance, 18*, 409-426.

Delgado, K. M. (2005). *Interactive effects of feedback type and feedback propensities on task performance.* Unpublished master's thesis, Wright State University, Ohio.

Diaz, D. D., & Sims, V. K. (2003). Augmenting virtual environments: the influence of spatial ability on learning from integrated displays. *High Ability Studies, 14*(2), 191-212.

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

Gilman, D. A. (1969). Comparison of several feedback methods for correcting errors by computer-assisted instruction. *Journal of Educational Psychology, 60*(6), 503–508.

Gonzalez, C. (2005). Decision support for real-time, dynamic decision-making tasks. *Organizational Behavior and Human Decision Processes, 96*, 142-154.

Goodman, J. S., & Wood, R. E. (2009). Faded versus increasing feedback, task variability trajectories, and transfer of training. *Human Performance, 22*, 64-85.

Goodman, J., Wood, R. E., & Hendrickx, M. (2004). Feedback specificity, exploration, and learning. *Journal of Applied Psychology, 89*, 248–262.

Graesser, A. C., Chipman, P., & King, B. G. (2008). Computer-mediated technologies. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. Driscoll (Eds.), *Handbook of research on educational communications and technology, 3rd edition* (pp. 211-224). New York: Lawrence Erlbaum Associates.

Guilford, J. P. (1956). The Guilford-Zimmerman aptitude survey. *The personnel and guidance journal, 35*, 219-224.

Guilford, J. P., & Zimmerman, E. S. (1948) The Guilford–Zimmerman aptitude survey. *Journal of Applied Psychology, 32*, 1, 24-35.

Guo, Y., Beck, J. E., & Heffernan, N. T. (2008). Trying to reduce bottom-out hinting: Will telling students how many hints they have left help? In B. P. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent Tutoring Systems: Lecture Notes in Computer Science, 5091* (pp. 774-778). Berlin: Springer-Verlag.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Hays, W. L. H. (1994) *Statistics* (5th edition). Belmont, CA: Wadsworth Group.

Hays, M. J., Lane, H. C., Auerbach, D., Core, M. G., Gomboc, D., & Rosenberg, M. (2009). Feedback specificity and the learning of intercultural communication skills. *Proceedings of the 14th International Conference on AI in Education*, 391-398.

Hoffman, R. R. (1998). How can expertise be defined?: Implications of research from cognitive psychology. In R.Williams, W. Faulkner, & J. Fleck (Eds.), *Exploring expertise* (pp. 81–100). New York: Macmillan.

ICT (2009). Intelligent guided experiential learning (IGEL): Tutoring for practice. Retrieved on March 10, 2010 from http://ict.usc.edu/projects/intelligent_guided_experiential_learning_tutoring_for_practice_igel_tp/

Jackson, G. (2007). *Investigating the roles of feedback categories within an intelligent tutoring system.* Doctoral dissertation, The University of Memphis, Tennessee. Retrieved August 14, 2009, from Dissertations & Theses: Full Text.(Publication No. AAT 3293764).

Johnson, W. L., & Rickel, J. W. (1996). Intelligent tutoring in virtual environment simulations. *ITS '96 Workshop on Simulation-Based Learning Technology*.

Jones, R. M., & Fleischman, E. S. (2001). Cascade explains and informs the utility of fading examples to problems. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society,* (pp. 459-464). Mahwah, NJ: Erlbaum.

Kalyuga, S. (2006). Instructing and testing for expertise: A cognitive load perspective. In A. M. Columbus (Ed.), *Advances in Psychology Research, Volume 46* (pp. 75-127). New York: Nova Science Publishers, Inc.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509-539.

Kalyuga, S. (2009). *Cognitive Load Factors in Instructional Design for Advanced Learners*. New York: Nova Science Publishers, Inc.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23-31.

Kennedy, R. S., Lane, N. E., Grizzard, M. C., Stanney, K. M., Kingdon, K., & Lanham, S. (2001). Use of a motion history questionnaire to predict simulator sickness. In *Proceedings of the sixth Driving Simulation Conference,* 79-89. Sophia-Antipolis (Nice), France: Neuf Associes.

Kenny, C., & Pahl, C. (2009). Intelligent and adaptive tutoring for active learning and training environments. *Interactive Learning Environments, 17*(2), 181-195.

Kester, L., & Kirschner, P. A. (2009). Effects of fading support on hypertext navigation and performance in student-centered e-learning environments. *Interactive Learning Environments, 17*(2), 165-179.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86.

Kjellin, M. S. (2005). "Areas of instruction" –A tool for assessing the quality of instructional situations. *Scandinavian Journal of Educational Research, 49*(2), 153-165.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279-308.

Lee, J. & Park, O. (2008). Adaptive instructional systems. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. Driscoll (Eds.), *Handbook of research on educational communications and technology, 3rd edition* (pp. 469-484). New York: Lawrence Erlbaum Associates.

Mangos, P. M. & Johnston, J. H. (2009). Performance measurement issues and guidelines for adaptive, simulation-based training. In D. A. Vincenzi, J. A. Wise, M. Mouloua, & P. A. Hancock (Eds.), *Human factors in simulation and training* (pp. 301-320). Boca Raton, FL: CRC Press.

Mason, B. J., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us.* Retrieved June 1, 2006, from http://dwb.unl.edu/Edit/MB/MasonBruning.html

McLaughlin, A. C., Rogers, W. A., & Fisk, A. D. (2008). Feedback support for training: Accounting for learner and task. *In Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*, 2057-2061.

Menaker, E., Coleman, S., Collins, J., & Murawski, M. (2006). *Harnessing Experiential Learning Theory to Achieve Warfighting Excellence.* Paper presented at the Interservice/Industry Training, Simulation & Education Conference, Orlando, FL.

Merrill, D. C., Reiser, B. J., Merrill, S. K., & Landes, S. (1995). Tutoring: Guided learning by doing. *Cognition and Instruction, 13*(3), 315-372.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97.

Mills, S. C., & Ragan, T. J. (1994). Adapting instruction to individual learner differences: A research paradigm for computer-based instruction. In *Proceedings of Selected Research and Development Presentations at the 1994 National Convention of the Association for Educational Communications and Technology Sponsored by the Research and Theory Division*, 525-546. Nashville, TN, Feb 16-20, 1994.

Moffat, S. D., Hampson, E., & Hatzipantelis, M. (1998). Navigation in a "virtual" maze: Sex differences and correlation with psychometric measures of spatial ability in humans. *Evolution and Human Behavior, 19,* 73-87.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*, 99–113.

Mory, E. H. (2004). Feedback research revisited. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745–783). Mahwah, NJ: Lawrence Erlbaum.

Munro, A. Fehling, M.R. and Towne, D.M. (1985). Instruction intrusiveness in dynamic simulation training. *Journal of Computer-Based Instruction*, *12*, 50-53.

Murray, T., & Arroyo, I. (2002). Toward measuring and maintaining the zone of proximal development in adaptive instructional systems. In *Proceedings of Intelligent Tutoring Systems: 6th International Conference, ITS 2002*, 749-758. Biarritz, France, and San Sebastian, Spain, June 2-7.

MYMIC LLC (1TPC7) (2004). *Outstanding issues in intelligent tutoring systems*. Retrieved on September 5, 2009 from http://www.google.com/url?sa=t&source=web&cd=2&ved=0CBwQFjAB&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.91.5335%26rep%3Drep1%26type%3Dpdf&ei=Qu0PTPD5KISdlgfi7emQDQ&usg=AFQjCNH7XQG71-jZx8Ny6NSLABPOXmZ5iw

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. Driscoll (Eds.), *Handbook of research on Educational communications and technology, 3rd edition* (pp. 125-144). New York: Lawrence Erlbaum Associates.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Ed.), *Instructional design for multimedia learning* (pp. 181–195). Munster, NY: Waxmann.

Nicholson, D., Fiore, S., Vogel-Walcutt, J. J., & Schatz, S. (2009). *Advancing the science of training in simulation-based training. Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting*, 1932-1934.

Oden, K. (2008). *Distributed team training: Effective team feedback.* Doctoral dissertation, University of Central Florida, Retrieved July 29, 2009, from Dissertations & Theses (Publication No. AAT 3340997).

Orvis, K. A., Horn, D. B., & Belanich, J. (2009). An examination of the role individual differences play in videogame-based training. *Military Psychology, 21,* 461-481.

Orvis, K. A., Horn, D. B., & Belanich, J. (2008). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior, 24*, 2415-2433.

Paas, F.G.W.C. (1992). Training strategies for attaining transfer of problem-solving skills in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429-434.

Pallant, J. (2007) *SPSS Survival Manual*. New York, NY: Open University Press.

Parasuraman, R., Mouloua, M., Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38*(4), 665-679.

Park, O. & Lee, J. (2004). Adaptive instructional systems. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 651–684). Mahwah, NJ: Lawrence Erlbaum.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences, 13*(3), 423-451.

Pine, A. (2009). Simulation: The (almost) real thing. *Proceedings Magazine, 135/12/1*(282). Retrieved from http://www.usni.org/magazines/proceedings/story.asp?STORY_ID=2119

Pridemore, D. R., & Klein, J. D. (1995). Control of practice and level of feedback in computer-based instruction. *Contemporary Educational Psychology, 20*, 444–450.

Priest, H. A., Durlach, P. J., & Billings, D. R. (in review). Developing collective training for small unmanned aerial systems employment: Workload and performance with multiple systems. ARI Technical Report.

Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist, 40*(1), 1-12.

Reisberg, D. (2006). *Cognition: Exploring the science of the mind, 3rd edition*. New York: W. W. Norton & Company.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13*(3), 273-304.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 15-22.

Salas, E., & Cannon-Bowers, J.A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*, 471-499.

Sanders, M. J. (2005). *The effect of immediate feedback and after-action reviews (AARs) on learning, retention and transfer.* Master's Thesis, University of Central Florida. Retrieved July 29, 2009, from Dissertations & Theses (Publication No. AAT 1428812).

Scerbo M. W., Freeman, F. G., Mikulka P. J., Parasuraman R., Nocero, F.D., & Prinzel, L. J., III (2001). The efficacy of psychophysiological measures for implementing adaptive technology. *Technical Report: NASA-2001-TP211018*. NASA Langley Technical Report Server.

Schatz, S. Bowers, C., & Nicholson, D. (2009). Advanced situated tutors: Design, philosophy, and a review of existing systems. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting*, 1944-1948.

Schooler, L. J., & Anderson, J. R. (1990). The disruptive potential of immediate feedback. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 702-708.

Schulze, K. G., Shelby, R. N., Treacy, D. J., Wintersgill, M. C., VanLehn, K., & Gertner, A. (2000). Andes: An intelligent tutor for classical physics. *The Journal of Electronic Publishing, 6*(1). Retrieved on July 2, 2010 from http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0006.110

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. *Interactive Learning Environments, 15*(1), 27-46.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

Shute, V. J., Woltz, D. J., & Regian, J. W. (1989). An investigation of learner differences in an ITS environment: There's no such thing as a free lunch. In D. Bierman, J. Breuker, & J. Sandberg (Eds.), *Artificial intelligence and education* (pp. 260-266). Amsterdam: IOS.

Shute, V. J. & Zapata-Rivera, D. (2007). Adaptive Technologies. In J.M.Spector, M. D. Merrill, J. van Merrienboer, & M. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology, 3rd Edition* (pp. 277-294). New York: Lawrence Erlbaum Associates.

Smits, M. H. S. B., Boon, J., Sluijsmans, D. M. A., & Van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: Effects on learning as a function of prior knowledge. *Interactive Learning Environments, 16*, 183-193.

Sweller, J. (2008). Human cognitive architecture. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. Driscoll (Eds.), *Handbook of research on educational communications and technology, 3rd edition* (pp. 369-382). New York: Lawrence Erlbaum Associates.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Tichon, J. (2007). Training cognitive skills in virtual reality: Measuring performance. *CyberPsychology & Behavior, 10*(2), 286-289.

Van Duyne, L. R. (2001*). Sequencing process and outcome feedback: Investigating the role of goal orientation during skill acquisition*. Doctoral dissertation, University of Central Florida. Retrieved July 20, 2009, from Dissertations & Theses (Publication No. AAT 3029054).

Van Duyne, L. R., Jentsch, K., Bowers, C., Burroughs, W., Cannon-Bowers, J., & Salas, E. (2001). Comparing the utility and sequencing of different types of feedback. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 1843-1847.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction, 18*, 211-222.

Van Merriënboer, J. J. G, & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147-177.

Waldman, A. (2009). *Serious gaming*. Retrieved on September 12, 2009 from http://www.ngaus.org/content.asp?bid=14851

Waldrop, P. B., Justen, J. E., & Adams, T. M. (1986). A comparison of three types of feedback in a computer-assisted instruction task. *Educational Technology*, *26*(11), 43–45.

Ward, P., Williams, A. M., & Hancock, P. A. (2006). Simulation for performance and training. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Feltovich (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 243-264). Cambridge: Cambridge University Press.

Widmayer, S. A. (2007). *Schema theory: An introduction*. Retrieved on June 4, 2009 from http://www.kihd.gmu.edu/immersion/knowledgebase/strategies/cognitivism/SchemaTheory.htm