Tailoring Instruction to the individual: Investigating the Utility of Trainee Aptitudes for use in Adaptive Training

Carla Landsberg
University of Central Florida

TAILORING INSTRUCTION TO THE INDIVIDUAL: INVESTIGATING THE UTILITY OF TRAINEE APTITUDES FOR USE IN ADAPTIVE TRAINING

by

CARLA R. LANDSBERG
B.S. University of Central Florida, 2007
M.A. University of Central Florida, 2012

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

Spring Term
2015

Major Professor: Mustapha Mouloua

ABSTRACT

Computer-based training has become more prolific as the military and private business enterprises search for more efficient ways to deliver training. However, some methods of computer-based training are not more effective than traditional classroom methods. One technique that may be able to approximate the most effective form of training, one-on-one tutoring, is Adaptive Training (AT). AT techniques use instruction that is tailored to the learner in some way, and can adjust different training parameters such as difficulty, feedback, pace, and delivery mode.

There are many ways to adapt training to the learner, and in this study I explored adapting the feedback provided to trainees based on spatial ability in line with Cognitive Load Theory (CLT). In line with the CLT expertise reversal effect literature I hypothesized that for a spatial task, higher ability trainees would perform better when they were given less feedback. Alternately, I hypothesized that lower ability trainees would perform better during training when they were given more support via feedback. This study also compared two different adaptation approaches. The first approach, called the ATI approach, adapts feedback based on a pre-measured ability. In this case, it was spatial ability. The second approach, called the Hybrid approach adapts initially based on ability, but then based on performance later in training. I hypothesized that participants who received Hybrid adaptive training would perform better.

The study employed a 2(spatial ability; high, low) X 2(feedback; matched, mismatched) X 2 (approach; ATI, Hybrid) between-subjects design in which participants were randomly assigned to one of the eight conditions. Ninety-two participants completed a submarine-based periscope operator task that was visual and spatial in nature.

The results of the study did not support the use of CLT-derived adaptation based on spatial ability; contrary to what was hypothesized, higher ability participants who received more feedback performed better than those who received less. Similarly, lower ability participants who received less feedback performed better than those who received more. While not significant, results suggested there may be some benefit to using the Hybrid approach, but more research is needed to determine the relative effectiveness of this approach.

ACKNOWLEDGMENTS

I would like to express appreciation to several individuals who have supported me throughout the process of completing this dissertation. Firstly I would like to thank my major advisor Dr. Mustapha Mouloua and Dr. Wendi Van Buskirk without whose guidance this dissertation would have been much more difficult if not impossible to complete. Dr. Mouloua kept me motivated without which I might never have finished. I have been lucky to learn from Dr. Wendi Van Buskirk, my co-worker but more importantly my mentor, who provided invaluable feedback throughout the entire process of this project. Work that I performed with Dr. Van Buskirk at the Naval Air Warfare Center Training Systems Division (NAWCTSD) provided the basis for my research ideas and her insights and support have aided me more than I will ever be able to express in words. I would also like to acknowledge my committee members Dr. Mark Neider and Dr. Clint Bowers who provided valuable constructive criticism throughout this process that helped improve my design and shape my thought processes.

I would also like to thank Naval Aviation Enterprise and the Office of Naval Research for funding this project under the Naval Innovative Science and Engineering (NISE) Workforce Development (WFD) and Independent Applied Research (IAR) programs. Additionally I am appreciative of the efforts Mr. Derek Tolley who helped create an experimental version of my testbed that allowed me to assess my research questions effectively and for putting up with me the various times I changed my mind. Thank you as well to several of my coworkers at NAWCTSD who provided me with feedback at various points in the course of this project. I appreciate the contributions of Dr. Gwendolyn Campbell, Dr. Randy Astwood, Dr. Cheryl Johnson, Alyssa Mercado, and Dr. Matthew Marraffino who helped me in various ways whether it was in shaping my design, helping me with data analysis, or providing sanity checks when I had been looking at a sentence for too long.

Lastly, I would like to express my deepest gratitude to my family for their support both during the several years it took me to finish this project, and in all of my academic pursuits that lead to this point. To my parents, who have always believed in me, and who have supported me both financially and emotionally- thank you. Observing your work ethics as I was growing up provided me with the motivation to succeed and made me the person I am today. I love you so much and this work was as much for you as it was for me. To my fiancé Skott, thank you for all of your emotional support during this process and for not running out the door when I was acting like crazy person due to stress.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

Problem Statement

Training is an important aspect of effective task performance, especially for tasks that are complex or those that have the potential to be dangerous. As such, billions of dollars are spent on training annually. A recent industry report estimated that training expenditures for United States-based corporations and academic institutions with over 100 employees totaled $59.7 billion dollars for 2011, an increase of 13% from 2010. The majority of this money ($31.3 billion) was spent on training staff payroll, i.e. personnel hired or assigned to perform training functions, while the rest of the total training budget ($9.1 billion) was spent on products and services from outside vendors and consultants (Training, 2011). Similarly, in the President's 2012 budget for the Department of Defense, $172 billion dollars of the Operations and Maintenance funds were allotted to the support of "training and readiness" in the military (U.S. Office of Management and Budget, 2012, pp.60). In light of the global economic recession, many companies are looking for ways to reduce costs and increase efficiency. Likewise, recent defense budget cuts have prompted the U.S. military to find ways to decrease spending. The figures above indicate that the majority of money spent on training in general is paid to instructors.

The challenge faced in both the private sector and in the military is how to reduce the cost of training while maintaining or improving training effectiveness. The literature suggests that one solution to this problem may be to increase the use of Computer-Based Training (CBT). This method of delivery generally does not require the same amount of instructor involvement as classroom-based training and has been shown to be effective when properly designed (Shearon, 2001). The literature on CBT, however, does not always conclude that it is the most effective method of delivery of training when compared to classroom-based or one-on-one tutoring (Kulik, & Kulik, 1982; Kulik, 1982; VanLehn, 2011). Moreover, studies that have compared human tutors to classroom lectures or computer-based training have shown that human tutors are generally more effective (Bloom, 1984; VanLehn, 2011). While there are many factors that may contribute to this deficit, it is possible that the design of some CBT systems lead to a decrease in effectiveness as compared to other types of training (Sitzmann, 2011). Producing training that can be presented via computer does not automatically make that training effective; designers must be careful to include empirically-tested principles of instructional design. As an example, some computer-based training is presented simply as scrolling text on a webpage; however research has shown that trainees perform better on CBT that allows them to take an active rather than passive role in the learning material (Brown & Ford, 2002; Bell & Kozlowski, 2008; Sitzmann, 2011). In one study, CBT was not shown to be more effective than other teaching methods when the training given was passive (Sitzmann, 2011).

In order to optimize the effectiveness of CBT, designers might consider incorporating Adaptive Training (AT), a type of CBT that tailors instruction to the needs of the learner and leverages some of the benefits of one-on-one tutoring. This method of instruction may be more effective than CBT alone because it can take into account trainee performance, individual differences, aptitudes, preferences, and/or personality factors (Kelley, 1969; Park & Lee, 2003; Shute, 2007). Vandewaetere, Desmet, and Clarebout (2011) point out that the consideration of learner characteristics is also an important aspect of instructional design. Unfortunately, despite a plethora of research in the area of adaptive training, there is still little agreement as to the best method of adaptation and there is little guidance for what type of adaptations are most beneficial for certain task types or learners (Vandewaetere et al., 2011).

Regarding adaptive training, the existing literature suggests that many considerations are necessary to create an effective system. In general, researchers agree that feedback is a necessary component of instructional design in order to increase learning and performance (Locke, 1968; Locke, Shaw, Saari & Lathem 1980; Locke & Latham, 1990). However, researchers also recognize that feedback does not universally improve performance (Kluger & DeNisi, 1996; Schmidt & Bjork, 1992). This is perhaps because there are several types of feedback (e.g. knowledge of results, process feedback, normative feedback, etc.) that can be presented to the learner and several ways to present it (Shute, 2007; Van Buskirk, 2011). Research seems to suggest that a well-designed AT system should provide some type of feedback to the learner while being mindful of the content and presentation of that feedback.

Moreover, it may be advantageous to consider the impact of individual difference variables when designing training. Previous research found relationships between the types of treatment a participant is given and aptitudes of the learner that may affect performance during training (see Berliner & Cahen, 1973; Cronbach & Snow, 1969; Snow & Lohman, 1984; Vandewaetere et al, 2011). Utilizing findings such as these, Cronbach (1957) suggested that by combining the correlational and experimental approaches, an optimal strategy for instruction could be found whereby it would be possible to pinpoint the best type of treatment for a particular type of person. That is, relating individual differences to training outcomes could aid in creating more effective training for different types of learners. However, it is not yet clear what individual difference variables are most effective for adaptive training interventions. In their review of the AT literature, Vandewaetere et al. (2011) revealed that despite extensive research that attempted to examine appropriate individual differences for use during training, only a few were successful. This may be due to concerns that arose during experimentation that either created confounds or low power issues.

First, when creating AT systems it is important to decide what the adaptation is going to be based on (i.e., performance, individual differences, affect), what to adapt (i.e., feedback, content, pace, sequence), and how to adapt it. As will be discussed below, there are several different adaptation approaches that can be employed. The objective of the current research is to compare two approaches of adaptive training using either individual differences (namely spatial ability) or individual differences and performance as the basis for adaptation. Through this research, I will also empirically examine some of the variables that affect the relationship between adaptive training manipulations and their effectiveness.

In the sections that follow, I will discuss the benefits of traditional CBT and Adaptive Computer Based Training (ACBT) and the use of feedback and individual differences in adaptive training. Finally, I will discuss how Cognitive Load Theory can be used as a theoretical framework to make decisions about what attributes of the learner can be used to create effective adapt training. I will also explain how recommendations from this approach differ from design recommendations based on a hybrid adaptive approach to adaptive training.

Computer-based Training

Recent advancement of computer technology has led to an increase in the use of CBT in classrooms, in the workforce, as well as in military domains (Bedwell & Salas, 2010; Kim, 2010; Lee, Owens, & Benson, 2002). CBT has numerous potential benefits over instructor-based classroom learning. A practical benefit of CBT includes lower overall costs than instructor-based training as a result of a decrease in delivery expenses and the capability to reuse the material several times (Kim, 2010). Additionally, CBT is readily accessible to trainees on their schedule and can be delivered in multiple locations at the same time (Lee, Owens, & Benson, 2002). Studies have shown that this type of training can be more effective and efficient than traditional classroom training; for example, a meta-analysis conducted by Kulik and Kulik (1991), which included 254 studies that used CBT, demonstrated that the use of CBT raised student test scores by .30 standard deviations in general and decreased overall training time as compared to conventional classroom teaching. Many studies have demonstrated additional benefits of CBT, showing evidence that CBT has the potential to increase knowledge retention (Sitzmann, 2011; Williams & Zahed, 1996), reduce training time (North, 1989), and increase motivation (Seyler, Holton, Bates, Burnett, & Carvalho, 1998; Sitzmann, 2011). Sitzmann (2011) examined the effectiveness of simulation-based training as measured by cognitive, affective, and performance outcomes using meta-analytic techniques. In this study, simulation training was defined as "instruction delivered via personal computer that immerses trainees in a decision-making exercise in an artificial environment in order to learn the consequences of their decisions" (p. 492). She found that the use of simulation CBT increased self-efficacy, acquisition of procedural and declarative knowledge, and retention as compared to groups that either received no training or comparative training that did not include simulation. However, there were not enough studies in the literature on trainee motivation, reactions to training, or performance on transfer tasks to include these variables in the analyses. Research that has shown CBT to be less effective than other methods of instruction (such as traditional classroom training or tutoring) may highlight the value of properly designed training. For instance, an analysis by Bedwell and Salas (2010) emphasized the importance of certain design considerations for CBT, such as matching delivery methods with the desired learning outcome and knowing when to use CBT versus classroom-based training. Likewise, Sitzmann (2011) found that simulation training was not more effective than other methods if the alternative instruction actively engaged learners and when simulation training was used as a substitute for alternative methods of instruction rather than a supplement.

CBT may be one way to approximate one of the most successful instructional methods: one-on-one tutoring. A study conducted by Bloom (1984) found that one-on-one tutoring was more effective by two standard deviations when compared to classroom-based learning. Similar research also found that one-on-one tutoring was more effective than classroom-based methods, but did not find as large an effect as was seen in Bloom's study (See Cohen, Kulik, & Kulik, 1982; Kulik, 1982; VanLehn, 2011). In a one-on-one setting, the instructor has the ability to adjust the material that is being presented to the trainee in real time based on the trainee's performance during training and their understanding of the material. The instructor can modify the delivery of the material, the pace of instruction, give additional examples, or provide detailed relevant feedback tailored to the trainee. However, Bloom (1984) also pointed out that one-on-one tutoring, while effective, would not be practical or fiscally responsible on large scale. He suggested the need for a training solution that would be as effective as individual tutoring but with less associated expense. He challenged educational researchers to discover the solution to what he called the 2-Sigma problem.

The literature suggested that both instructor-based training and computer-based training could offer benefits, however the evidence has not decisively named a frontrunner for superior training. As an example, in a study that compared instructor based learning to CBT, participants who received the instructor-based learning reported higher satisfaction and transfer of knowledge; however participants who received CBT retained more knowledge over a 60 day period. Additionally, both methods lead to significant learning gains from a pre-test to post-test (Kim, 2010). While both methods showed benefits, these results suggest that more research will be required in order to solve the 2-Sigma problem.

Adaptive Computer-Based Training

The evidence presented above suggests that perhaps combining the benefits of one-on-one tutoring and CBT may improve overall training effectiveness and efficiency while decreasing the cost and time associated with training. One method that may bridge the gap between the benefits of one-on-one tutoring and computer based training is called Adaptive Computer-Based Training (ACBT), or Adaptive Training (AT). Adaptive training can generally be described as training that matches a student’s individual differences in either aptitude or performance to the instructional techniques they receive during training (Park & Lee, 2003). In other words, training content, sequence, pace, or difficulty could be altered in order to match student ability, preference, motivational state or other characteristics. More specifically, AT is “training interventions whose content can be tailored to an individual learner’s aptitudes, learning preferences, or styles prior to training and that can be adjusted, either in real time or at the end of a training session, to reflect the learner’s on-task performance” (Landsberg, Van Buskirk, Astwood, Mercado, & Aakre, 2011, pp. 9). These definitions are broad and imply that AT can be applied in any number of ways; there are many aspects of training that can be tailored, and there are many characteristics that can be used as the basis for adaptation. Perhaps because of this, despite 50 years of research on AT (both pencil-and-paper and computer-based) there has been little agreement on the best methods for tailoring instruction to individual learners (Shute, 2007; Vandewaetere et al., 2011).

Research has indicated that adaptive training can be effective in general when compared to non-adaptive training (e.g., Bauer, Brusso, & Orvis, 2012; Corbalan, Kester, & van Merrienboer, 2008; Graesser, Conley & Olney, 2012), however the empirical question still remains: how best to apply the adaptation? Many approaches to adapting an instructional strategy have been studied, including adaptation of difficulty, feedback, and pacing, sequencing and modality of instruction. Decisions regarding instructional interventions can be complicated because different instructional adaptations may be better (or worse) depending on learner individual differences or the type of training task being learned. For example, research has shown that process feedback may be more useful than outcome feedback in situations where decision making is being trained while highlighting might be more effective than process feedback for procedural tasks (Bangert-Drowns, Kulik, Kulik & Morgan, 1991; Buff & Campbell, 2002; Van Buskirk et al., 2009).

Types of Adaptive Training

There are several methods that can be used to adapt training. The oldest and most widely used of the methods is called macro-adaptation. In this method, the adaptation of training occurs on a broader level, allowing the instructor to choose from a finite set of instructional changes such as adjusting the learning goals or the depth or presentation of content (i.e., visually, verbally, or auditory). These instructional changes may be based on general trainee dimensions such as general ability, learning style, or achievement levels in the curriculum (Mödritscher, Garcia-Barrios, & Gütl, 2004). For this method, all of the adaptation occurs prior to the start of training. Implementation of this method of AT, although widespread, has been unsystematic and very few studies have performed effectiveness evaluations (Park & Lee, 2003; Vandewaetere et al., 2011). Other methods include the micro-adaptive method, the collaborative constructivist approach, the Aptitude Treatment Interaction (ATI) method, and the two-step method/ hybrid method.

The micro-adaptive method is used to adapt instruction to the trainee based on their real time performance during the training task. Measures of performance can include variables such as response errors, reaction times, or measures of workload during training (Kelley, 1969; Mödritscher et al., 2004). Intelligent Tutoring Systems (ITS) fall into this category of adaptive training. ITS employ artificial intelligence in an attempt to provide training that is more similar to what would be provided by a human tutor. ITS generally include four components: a problem solving environment where students encounter the learning material, a model of domain knowledge that student answers are compared against, a student model that maintains information about the student's current understanding of the material, and a pedagogical module that dictates how instruction is delivered to the student based on the student model (Corbett, Koedinger, & Anderson, 1997). In the first 20 years or so of study of these systems, much of the research in this field was focused on improving the artificial intelligence algorithms rather than verifying their educational benefits or verifying their training effectiveness (Corbett, Koedinger, & Anderson, 1997). However, more recently, several ITS have been evaluated for their instructional effectiveness with success in creating learning gains when compared to control groups (Corbett, 2001; VanLehn, et al., 2007). While these systems have been shown to be effective, possibly because they most closely approximate one-on-one tutoring (Park & Lee, 2003; Stottler & Vinkavich, 2006), micro adaptive training can be more expensive and time consuming to develop than other methods, such as the macro adaptive and ATI methods (Corbett, et al., 1997; Pew & Mavor, 1998; Shute & Zapata-Rivera, 2008). Additionally, different ITS architectures have been used to teach material in the same domain (Graesser, Conley, & Olney, 2012), and research has not been carried out that would suggest which architecture is best. This may depend on learner characteristics and individual differences related to the material being trained.

The constructivistic-collaborative approach is based on constructivist and collaborative theories of learning that encourage the learner to take a more active role in their education. This approach focuses more on the process of learning than the learning outcomes themselves (Akhras & Self, 2000). In this type of training, the learner builds their understanding of the material through their interactions with the domain content in the context that the material will be applied. By interacting with the information, the learners build their own mental model rather than being explicitly given the important concepts by the trainer. Collaborative-constructivism stresses the importance of interacting with other learners to gain a shared understanding of material that serves to solidify the learners' mental models (Du & Wagner, 2005). Because the constructive theory of learning focuses more on the learning process than the outcomes, traditional ITS architectures are not ideal for this approach and different types of systems are necessary (Akhras & Self, 2000).

The fourth method that has been used to adapt training, and which will be one of the focuses of this study is called the Aptitude-Treatment Interaction (ATI) approach. In this approach, instruction is tailored to specifically match trainee aptitudes such as learning or cognitive styles, prior knowledge, personality, or ability. Cronbach and Snow (1977) suggested that optimal learning occurs when instruction is matched with trainees' individual differences such that a trainee with certain aptitudes would benefit maximally from training of one type, while a trainee with different aptitudes would benefit from training of a different type. Studies have examined several individual difference variables that have been related to performance such as intellectual ability (Snow & Lohman, 1984), cognitive and learning styles/ preferences (Angeli, Valanides & Kirschner, 2009; Davis, 1991; Massa & Mayer, 2006; Messick 1994; Snow & Lohman, 1984), prior knowledge (Glaser & Nitko, 1971; Kalyuga, 2005; Shin, Schallert, & Saveyne, 1994), anxiety (Deutsch & Tobias, 1980), personality (Barrick & Mount, 1991; Bauer, Brusso & Orvis, 2012; Deutsch & Tobias, 1980; Herold, Davis, Fedor & Parsons, 2002; Horak & Horak, 1982; Yeany, Dost, & Matthews, 1980), achievement motivation (Linn, 1993;Wolf & Smith, 1995), and self-efficacy (Park & Lee, 2003; Sitzmann, 2011). Research indicates different levels of these individual differences can alter the effectiveness of adaptive training. Some examples will be discussed in later sections.

Yet another method of adaptive training combines the ATI method with the micro method. This is called the hybrid or two-step method of adaptive training (Park & Lee, 2003; Tennyson & Rothen, 1977; Tennyson & Christensen, 1988) and will also be explored as part of this research. In this approach, initial conditions of instruction are established based on the ATI method of AT, while later in training real-time evaluation of the learner's performance (i.e., a micro-adaptive approach) is used to adjust training parameters. The hypothesis for this approach is that pre-task aptitudes would be more predictive of learner performance in the beginning of training, while on-task performance would be more predictive later in training (See Figure 1). Research demonstrated that there are some aptitude variables, such as cognitive ability and prior knowledge that are valuable for placing learners into initial instructional treatments (Park & Lee, 2003). However, studies by Park and Tennyson (1980; 1986) demonstrated that the predictive value of these variables decreased over the course of training, possibly due to the effects of other variables' interactions with the chosen aptitude variables. A review of the adaptive training literature illustrated that there was a lack of evaluation of two-step approach (Landsberg et al, 2011). Additionally, in order to employ this method effectively, ATIs that are effective for training must be identified empirically.

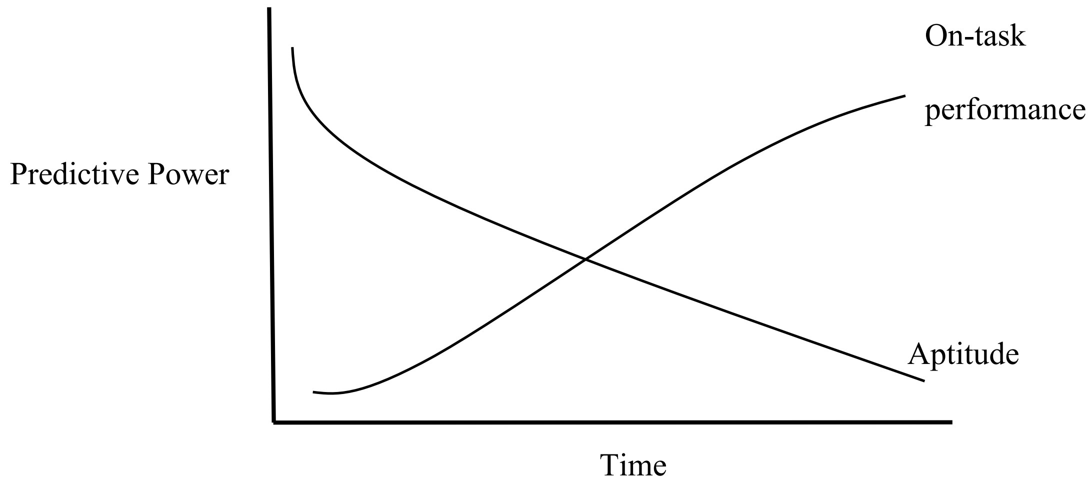


Figure 1: Hypothesized relationship between the predictive power of performance and aptitude variables over time (adapted from Park & Lee, 2003).

Adaptive Training Effectiveness

<u>Individual Differences</u>

Several different individual difference variables may be important to consider for their relationship to learning. One of these, spatial ability, is a component of general intelligence and can broadly be defined as a person's "skill in representing, transforming, and recalling symbolic, nonlinguistic information" (Linn & Petersen, 1995, p. 1482). Table 1 shows that spatial ability is not a unidimensional construct, but rather it is separable into several factors (Carroll, 1993; Kozhevnikov & Hegarty; 2001, Lohman 1988; McGee, 1979). Three factors that are discussed in the literature are spatial visualization, spatial orientation, and spatial relations (Kozhevnikov & Hegarty; 2001, Lohman 1988; McGee, 1979). Spatial visualization is the ability to manipulate (rotate, twist or invert) objects without reference to oneself. This has also been called object-based perspective transformation (Kozhevnikov & Hegarty, 2001; Kozhevnikov, Motes, Rasch,

& Blajenkova, 2006; Lohman, 1988; McGee, 1979) and it can be measured using paper folding tasks, mental rotation tasks, and form board tests. The second factor, spatial orientation or egocentric perspective transformation, is the ability to imagine how a visual array would look if seen from a different perspective relative to the observer (Erkstrom, French, & Harmon, 1976; Kozhevnikov & Hegarty, 2001; Kozhevnikov et al., 2006; McGee, 1979). This type of spatial ability can be measured with perspective taking tasks such as the object or map perspective test or the Guilford-Zimmerman Spatial Orientation task (Kozhevnikov & Hegarty, 2001).   Lastly, the spatial relations ability is the speeded mental rotation of two dimensional items (Lohman, 1988; Kozhevnikov & Hegarty, 2001). This ability can be measured using card rotation or cube comparisons.

Table 1. Measures of Spatial Ability by Factor

Factor	Measure	Description
Spatial Visualization	Paper Folding Test (Erkstrom et al., 1976)	Each item shows drawings of two or three folds made in a sheet of paper. The last drawing shows a hole punched in the folded paper. The participant must select one of five drawings that shows how the punched sheet would appear if opened
	Mental Rotation Test (Vandenberg & Kuse,1978)	For each item, participants are shown a 3D figure and are asked choose which two of the four comparator figures is the same object but rotated around an axis
	Form Board Task (Erkstrom et al., 1976)	For each item, participants are presented with five figures that can be arranged to form an object. The participant has to indicate which pieces can be put together to form the shape given for that item
Spatial Orientation	Object Perspective Task (Koheznikov & Hegarty, 2001)	Participants view an array of objects and are asked to imagine being at the position of one of them. They are asked to "point" in the direction of a second object in the array by drawing a line on an answer sheet from their imagined station to their imagined heading
	Map Perspective Test (Koheznikov & Hegarty, 2001)	Participants view a map that has five landmarks and are asked to imagine being at the position of one of them. They are asked to "point" in the direction of a second object in the array by drawing a line on an answer sheet from their imagined station to their imagined heading
	Guilford-Zimmerman Spatial Orientation task (Guilford & Zimmerman, 1948)	Each item shows two pictures of the bow of a ship facing the shore. Participants are asked to judge what direction the ship has moved from the first picture to the second
Spatial Relations	Card Rotation Task (Erkstrom et al., 1976)	Participant judges which of the five 2D figures is a rotation of the target image as quickly and accurately as possible
	Cube Comparison Task (Erkstrom et al., 1976)	Participants are presented with drawings of two cubes that have letters and numbers printed on their sides. Participants judge if the two cubes are the same or different

Fairly reliable differences in spatial ability have been demonstrated between sexes when measured by some spatial tasks. For instance, when measured using a mental rotation task, males tend to exhibit higher spatial ability than females; a finding that is stable between different ages and cultures (Linn & Petersen, 1985; Peters et al., 1995). Not all spatial tasks produce sex differences, however (Alyman & Peters, 1993; Peters et al., 1995). In general, it seems that gender differences are more reliably produced on Vandenberg and Kuse's (1978) Mental Rotation Test (MRT) than other measures of spatial visualization such as Erkstrom et al. (1976) Card Rotation Test or the Picture Folding Test (Peters et al., 1995).

Research has suggested that male and female differences on tests of spatial ability may be partially explained by differential experience on spatial tasks (Baenninger & Newcombe, 1989; De Lisi & Cammarano, 1996). For example, males were more likely to have experience with tasks that are more spatially oriented, such as playing basketball or videogames than females and this may help to explain why they perform better on measures of spatial ability. Recent research also demonstrated that playing video games leads to improvements in cognitive functions related to spatial ability and can lead to improved performance on spatial tasks- such the MRT. This effect was demonstrated with a type of action video game that is generally more appealing to males than females (Spence & Feng, 2010). A meta-analysis performed by Baenninger and Newcombe (1989) found a weak but reliable relationship between spatial experiences and sex. As part of their analysis, Baenninger and Newcombe (1989) also looked at the relationship between practice and performance on tests of spatial ability. They found that both sexes showed a similar pattern of improvement on spatial tests following practice. This research suggests that sex differences on spatial measures should be smaller between males and females with similar experience on spatial tasks (such as experience with videogames). Spatial ability is an important aspect of performing many tasks; for example, performing laparoscopic surgery, 3D modeling, calling angle on the bow, operational planning and navigating, and map reading. Recently, there has been an increase in the use and study of game-based training, simulation-based training, and scenario-based training- particularly in the military domain. This type of training tends to be spatially oriented and because of this, spatial ability is the individual difference variable that has been chosen as the focus of this study.

Research has also demonstrated a relationship between working memory capacity and performance on certain types of tasks. Working memory has been defined as a brain system that is responsible for the temporary holding and manipulation of information (Baddeley & Hitch, 1974). One of the most widely accepted models of working memory was developed by Baddeley and Hitch in 1974 and later modified by Baddeley in 2000 to include the episodic buffer. In Baddeley's model, working memory is composed of four components: the phonological loop, the visuospatial sketchpad, the central executive, and the episodic buffer (See Figure 2). The phonological loop, made up of a storage component and a rehearsal component, is theorized to hold verbal and auditory information. In contrast, the visuospatial sketchpad is made up components that hold visual, spatial and kinesthetic information. The third component, the episodic buffer, is a limited capacity storage system that connects the activities phonological loop and the visuospatial sketchpad to the long term memory system. All of these systems are controlled by the central executive, which is a limited capacity attentional control system that allocates resources and coordinates activities between the other components.

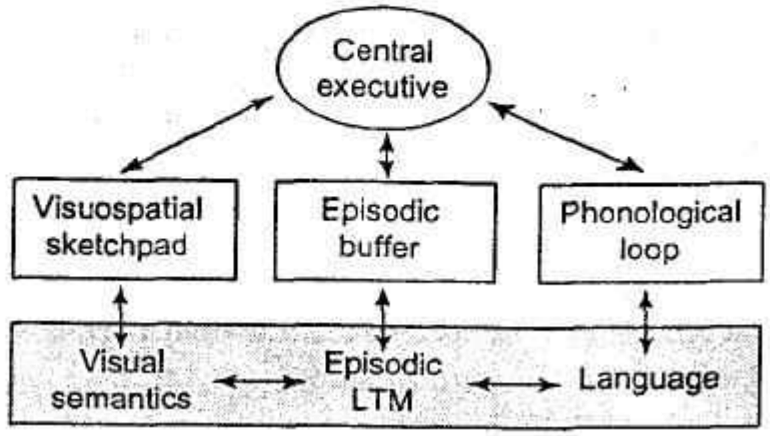


Figure 2. The Baddeley Model of Working Memory (Baddeley, 2000).

Several studies have also indicated that trainees who were higher in working memory capacity would perform better when given multimedia training than trainees who had lower working memory capacities. For instance, in a study performed by Batka and Peterson (2005), the researchers concluded that participants who were measured to have higher working memory capacity performed better in general on transfer questions after receiving multimedia training that explained the formation of hail. They also found that trainees with lower working memory capacity performed better when the multimedia training they received presented animation and narration simultaneously rather than sequentially. In an AT study performed by Corbalan, Kester, and van Merriënboer (2008), working memory, along with either shared or program control of task selection, was used as the basis for selecting difficulty and support levels in a dietetics training task. When compared to non-adaptive groups, participants in the adaptive groups learned more efficiently, that is they earned higher scores with less reported mental effort, and reported higher task involvement.

An interesting finding in the literature indicated a relationship between working memory capacity and spatial ability. Specifically, in a study performed by Just and Carpenter (1985), when participants determined to be low spatial ability (as determined by a test of spatial visualization) performed a cube comparison task, eye tracking determined they required more rotations of a cube face than participants with high spatial ability. The authors concluded that the low spatial ability participants had forgotten an intermediate representation of the image during rotation because they had difficulty simultaneously keeping the image in memory while performing the rotation. They suggested that this indicated that individual differences in working memory capacity may be one of the underlying explanations for differences in spatial ability (Shah & Miyake, 1996)

Recently, a number of researchers (Park and Lee, 2003; Pashler, McDaniel, Rohrer & Bjork, 2009) concluded that despite extensive research on the ATI approach, consensus has not been found that the use of this approach significantly improved training effectiveness. Researchers have also exposed several limitations of the ATI approach (Park & Lee, 2003). Some of these issues include the problem that the levels of abilities that are chosen prior to training may change during training, making them less effective later in training. Additionally, ATIs that are identified for a particular task and domain area may not generalize to other tasks or domains. Lastly, ATIs that are identified in experiments that take place in laboratory settings may not generalize to real-world settings. In a recent review, Pashler and colleagues (2009) assessed the effectiveness of a student preference called learning style for use in making adaptive training decisions. They found that although extensive research continues on the concept, which proposes that certain modes of instruction are more effective for particular individuals (i.e.

pictures for visualizers and text for verbalizers), there is little empirical evidence to support the incorporation of this variable into instruction.

Although the research cited above did not support the use of the ATI approach, several studies found that it can be effectively implemented. For instance, studies performed by Park and Tennyson (1980) and Tennyson and Rothen (1977) have shown that using pre-task aptitude measures to place trainees into instructional conditions can beneficial; when trainees' aptitudes were used to determine the amount of training they would receive, they took less time to finish the training and also performed better on a post-test than trainees who received the same amount of instruction, but that was not based on their aptitude scores. Additionally, there are aptitude measures that have been shown to be effective for placing learners into instructional treatments (Park & Lee, 2003). Moreover, other studies have shown support for using certain aptitudes as adaptation variables. A pencil-and-paper based study by Snow and Lohman (1984), found that students with lower intellectual ability benefited from more structured and less complex training while students higher intellectual ability performed better when given less structured and more complex training. Further, in the review performed by Pashler et al. (2009), the authors asserted that many studies they reviewed were not properly designed to answer the research questions that were posed. Specifically, in order to find the ATI approach to be effective, several conditions must be met: Firstly, participants must be tested prior to training for levels of a specific aptitude, then participants must be randomly assigned to one of two or more training conditions (one condition that is adapted to that aptitude and one that is not), and lastly on a post-test, students who were assigned to the training that was matched to their ability should outscore students who were assigned to the training not matched to their ability. In the review by Pashler et al. (2009), the authors noted that only one of the papers they reviewed showed this type of interaction, and this study focused on learning preference, not ability. In this study, I plan to execute a properly designed study in order to show that the ATI approach to adaptive training can be effective when spatial ability is used as the adaptation variable. I will also examine several possible mediators and moderators of this effect.

Feedback

Feedback is information that is given to the learner regarding some aspect of their performance on a task (Ilgen, Fisher & Taylor, 1979; Kluger & DeNisi, 1996). This broad definition can encompass many different types of feedback: outcome feedback, process feedback, normative feedback, progress or velocity feedback, to name a few. Outcome feedback is the simplest form of feedback and provides the trainee information about their performance on a task (Kluger & DeNisi, 1996). This might be a score, a percentage correct, or a message that informs the trainee if they were correct or incorrect. Next, process feedback gives the trainee information on how to perform the task (Kluger & DeNisi, 1996). This type of feedback goes beyond telling the trainee how well they performed the task, and tells them how to perform the task better in the future. An example of process feedback for a driving task might be, "Don't forget to check your rearview mirror before you reverse". Another type of feedback, normative feedback, compares the trainee's performance on a task to the performance of others on the same task (Smither, Wohlers, & London, 1995). This allows the trainee to determine how well they are doing on a task compared to their peers. For instance, the feedback might say, "You are in the 70^{th} percentile for this task, 30% of your peers performed better than you." Lastly, velocity feedback compares the trainee's performance to their own performance on a task over time, allowing the trainee to estimate how rapidly they are progressing (Kozlowski et al., 2001). As an example, the feedback might say, "You scored 30% higher on this scenario than the last scenario you performed". Some of these types of feedback may be more useful than others when trying to train certain types of tasks. Studies have also demonstrated that different types of feedback may be more beneficial depending on a trainees' cognitive ability.

Several studies have suggested that process feedback is beneficial for training, particularly when the task is complex. Buff and Campbell (2002) compared presentation of process and outcome feedback during a study and found that trainees who received process feedback scored significantly higher on a post-test than participants who received either outcome or no feedback during training on a decision making task. Further, in a study where process, normative, outcome, and no feedback were compared, researchers found that participants who received process feedback outperformed participants who received any other type of feedback on a visual decision making task (Astwood, Van Buskirk, Cornejo & Dalton, 2008). Because process feedback focuses on the behaviors that are necessary to perform a task well rather than the outcomes of those behaviors, process feedback might have a larger impact on training performance than presentation of outcome feedback (Earley, Northcraft, Lee, & Lituchy, 1990).

Kelley and McLaughlin (2012) examined the relationship between individual differences in cognitive abilities and the type of feedback (they referred to this as the amount of task support) presented during training. In the study, they used two populations thought to be higher in different cognitive abilities (crystallized intelligence and fluid intelligence) and a cuing task that required the use of one of these abilities or the other. Kelley and McLaughlin found that participants who were higher in crystallized intelligence performed more accurately and quickly on a task when they were given outcome feedback only on the task that required the use of crystallized intelligence, but performed better on the task that required fluid intelligence if they were given both outcome and process feedback. This finding supports the notion that those who are higher in an ability that relates to the task being trained require less feedback than trainees who are lower in that ability. This research mirrors a concept called the expertise reversal effect where instructional methods that work well for novices might actually be harmful for those with more knowledge in the domain (Kalyuga, 2005). The expertise reversal effect concept was developed within the framework of the Cognitive Load Theory, which will be discussed below.

Cognitive Load Theory

Cognitive Load Theory (CLT) can be used to support adaptive training design decisions. According to this theory, there are three types of cognitive load associated with training: intrinsic, germane, and extraneous. Intrinsic load can be described as the cognitive processing that is necessary to understand the instructional material; the degree of this type of load depends on task difficulty or the nature of the task. Germane load is the cognitive processing that is necessary to relate new information to prior knowledge and organize it into schemas (i.e. learning). Extraneous load is the superfluous cognitive load that fluctuates based on instructional design (DeLeeuw & Mayer, 2008; Sweller, 1988); a well-designed instructional system will keep this type of load low, while a poorly designed system will cause this load to increase. These three types of load all compete for a limited amount of working memory resources, and the total amount of these loads cannot exceed the total working memory capacity (Paas, Renkl, &

Sweller, 2003). The goal of CLT-based training is to increase germane load and decrease extraneous load while encouraging the learner to create schemas for learning objectives. In essence, creating schemas is the process of moving information from working memory (a limited capacity system), to long term memory (a system that is nearly unlimited). Cognitive load is lowest when automated schemas are being used (Tuovinen & Sweller, 1999). Early in training, more working memory resources are being utilized to understand the material and relate it to past experiences and knowledge; however, once learning material becomes schematized it requires little working memory so that the majority of resources can be used to integrate new information. In practice, research has illustrated that worked examples, a strategy supported by CLT, may be beneficial for students with low prior knowledge while problem solving may be better for learning for a student with high prior knowledge (Schaefer & Dyer 2011). Rey and Bushwald (2011) found that novice trainees who received more information while learning a task performed better than novice trainees who did not receive extra information; interestingly this result was reversed for trainees who had more knowledge of the task before training. In addition, the novice trainees who received extra information reported lower levels of cognitive load than those who did not. The opposite was found for the more experienced trainees. The results of Rey and Buchwald (2011) supported the existence of the expertise reversal effect. Similarly, research by Tuovinen and Sweller (1999) reported that instructional strategies supported by CLT were more effective for trainees with low domain knowledge than those who are more experienced.

Until recently, CLT has been largely applied to "well-structured procedural and conceptual domains" or classroom-based material (van Merriënboer & Sweller, 2005 pp. 156) and is only now being applied to more complex real-life tasks. Because these types of tasks have more interrelated elements and impose a higher level of intrinsic load, it is not always possible to decrease the cognitive load to a manageable level by decreasing extraneous load alone (van Merriënboer & Sweller, 2005). Including AT can help by keeping cognitive load at a manageable level by diminishing extraneous load during training.

Cognitive Load Theory and Adaptive Training

Theoretically, the use of adaptive training should ease the burden on working memory resources and encourage germane processing because it is suited to the individual needs and ability of the learner. In the case of the ATI approach being explored in this research, students with low ability will receive extra support to augment their cognitive resources and encourage germane processing while students with higher ability receive less support decreasing extraneous load and allowing them to use their available cognitive resources to integrate new information.

The effectiveness of the hybrid method of adaptive training may be harder to support using CLT. For instance, one design principal based on CLT suggests trainees should receive more detailed feedback or a higher level of scaffolding in the beginning of training a novel task followed by a decrease in the amount of support as trainees become more experienced. One example given in the literature is giving novice trainees worked examples, partially worked examples, and then full problems, in other words stepping down the level of instructional support as training progresses (van Merriënboer & Sweller, 2005). As mentioned previously, in the hybrid approach, trainees receive initial instruction based on pre-task measures similar to training designed using the ATI approach and so training is matched to trainee ability. However, later in training, the basis of adaptation changes and instructional decisions are based on a micro adaptive approach, i.e. based on the trainees' current on-task performance. This would suggest that some trainees would receive more support later in training, while some would receive less- depending on their performance in initial training trials. In other words, for some participants feedback might actually increase later in training, or it is possible that some participants would receive less support early in training. Additionally, CLT does not necessarily suggest when feedback support should decrease and this drop-off might differ between high and low ability trainees. In the ATI method of adapting feedback, the amount of feedback provided remains stable throughout training based on ability alone, while in the hybrid approach to AT, this design decision is also based on task performance or level of ability. CLT does not offer guidance on whether it is better to base design decisions on aptitude or performance and this empirical question will be explored in this research.

Gaps in the Current Adaptive Training Literature

The review of current research above suggests that there are several gaps in the literature regarding adaptive training. Firstly, while AT has been demonstrated to be effective, little is known about what it is that makes it so. This issue is complicated by the sheer number of ways it is possible to effect that adaptation. More specifically, there are numerous approaches that can be used to adapt training (e.g., micro, ATI, hybrid), a large number of ways to tailor instruction, such as manipulating difficulty, pace, content, etc., and several individual difference and/or performance variables that can be used as the basis of adaptation. Moreover, the current literature base does not seem to be in agreement as to the best method of adapting training, and very few studies report training effectiveness evaluations that are necessary to make this determination. Additionally, even fewer studies make direct comparisons of different methods of adaption, often comparing adaptive training to non-adaptive training. Research has also not identified individual difference variables that are useful for ATI style training approaches. While some variables have been related to performance consistently (e.g., working memory, spatial ability, motivation), the ATI method of adaptive training has not received strong support. As mentioned previously, this may be due to a lack of experimental rigor in this area of research. An overarching model of AT that integrates individual differences (e.g., motivation, personality, working memory, etc.) does not yet exist to guide designers during development of AT training.

While CLT can be used to support design decisions in AT, there are still questions as to which methods are more effective for learning and what criteria should be used to make adaptive decisions. For example, CLT does not provide recommendations for choosing between performance and ability as adaptive variables. In addition, while CLT has been applied extensively in classroom and well-structured settings, it has just begun to gain popularity for use in more complex and realistic training settings. More research is needed to determine that CLT principles and guidelines hold true for multifaceted tasks and if AT can help ease the burden of the extra cognitive load burden they impose. This goal of this dissertation is to examine some of these questions empirically and to help create an empirical basis to support AT decisions.

CHAPTER TWO: CURRENT STUDY

Purpose

The current study will examine the efficacy of using an aptitude, spatial ability, as an adaptation variable. Spatial ability will be used to determine the amount of feedback that is provided to participants during training on a spatially-oriented task. Analyses are expected to show that matching feedback to the participants' spatial ability leads to better performance than when spatial ability is not matched. In addition, spatial ability as an adaptation variable will be studied in the context of two different approaches of adaptive training: an ATI approach and a Hybrid approach. This study will be the first to compare these two methods of adaptive training empirically. While there is a lack of research on the hybrid method of adaptive training in general (see discussion above), it is expected that participants who receive this type of training will perform better than those who receive ATI training. More specifically, I expect that the matched hybrid group will outperform all other training conditions. In addition, as part of this research, several individual difference variables will be explored in order to determine elements that contribute to adaptive training effectiveness. Variables explored will include self-efficacy, frustration, and working memory capacity.

Hypotheses

In line with cognitive load theory, lower ability participants who receive more feedback should experience a decrease in the amount of extraneous load allowing for an increase in capacity for germane processing; this will allow these participants to spend more cognitive resources on learning the task. On the other hand, consistent with the expertise reversal effect, providing trainees who are already high in spatial ability with more feedback may increase extraneous load and interfere with learning. In other words, people with low spatial ability may require more help during training to offset their detriment, while people with higher spatial ability would be hindered by extra help that they do not need- i.e. it would require someone with higher spatial ability to use more resources to ignore extra material, or it would require them extra resources to recognize that the extra help is not helpful. As such, I would expect that participants of both low and high spatial ability will perform better in matched conditions because they will be receiving the appropriate amount of feedback for their respective ability levels, allowing them to utilize additional resources for germane processing rather than extraneous processing. This should also allow participants in matched conditions to create more effective schemas and allow them to perform the task faster.

I also expect that participants who are matched on their ability will receive higher scores during transfer. According to CLT, groups that are ability matched should experience more germane processing leading to better transfer of training (DeLeeuw & Mayer, 2008). In fact, according to DeLeeuw and Mayer (2008), transfer performance is the best measure of germane load.

Lastly, I expect that participants in matched conditions will benefit from reduced amounts of extraneous load. I believe that the reduction in load that occurs during training will allow participants in matched conditions to maximize their learning time and achieve higher learning gains from pre- to post-test. Therefore, I hypothesize that:

Hypothesis 1: Participants in matched feedback conditions will perform better than those in mismatched feedback conditions

> *Prediction a: High spatial ability trainees who are given less feedback (matched) will perform better than high spatial ability trainees who are given more feedback (mismatched).*
>
> *Prediction b: Low spatial ability trainees who are given more feedback (matched) will perform better than low spatial ability trainees who are given less feedback (mismatched)*

No studies in the literature have compared the hybrid and ATI approaches to adaptive training directly. In fact there is very little research on the two-step/hybrid method and research has been conflicted at best for the ATI method of AT. As mentioned earlier, CLT can be used as a framework to support the use of adaptive training, because matching instruction to a trainees ability levels or performance should reduce extraneous cognitive load. CLT does not yet help instructional designers choose between different methods of AT. However, participants whose training is based on both ability and performance may have gotten training that better matched their current level of ability. According to Park and Tennyson (1980; 1986), ability variables measured pre-training may decrease in effectiveness over time. If this is the case, ability measures would be more useful in the beginning of training when there is no performance data. However, later in training, on-task performance may be more predictive of future performance on the task and a better candidate as the basis of adaptation decisions. If hybrid matched participants' training is better matched to their current level of ability, it will lead to a decrease in extraneous load, allowing more resources for germane processing. During the post-test, participants in the hybrid groups should perform better because they created more effective schemas during training. Additionally, participants in hybrid matched conditions should experience the lowest extraneous load and achieve higher learning gains than all of the other instructional conditions. As such, I hypothesize that:

> *Hypothesis 2: Participants who receive hybrid adaptive training will perform better than those who receive ATI adaptive training*
>
> > *Prediction a: High ability participants who receive hybrid training will perform better than high ability participants who receive ATI training*
> >
> > *Prediction b: Lower ability participants who receive hybrid training will perform better than lower ability participants who receive ATI training*
> >
> > *Prediction c: Participants who receive hybrid matched training will do better than all other groups*

As mentioned above, the MRT will be used to measure participant spatial ability. Based on the research cited above, I additionally hypothesize that:

> *Hypothesis 3: Males will score higher than females on the Mental Rotation Test with experience on spatial tasks as measured by demographic experience questions mediating the relationship between gender and scores on the MRT and on the experimental task*

In addition to improving post-test performance and overall learning gain, I predict that ability matched training and hybrid training will improve performance over the course of actual training. This improvement would occur due to the reduction of extraneous load that participants experience while training, causing the matched and hybrid groups to master the task faster than participants in other groups. Therefore, I hypothesize the following:

> *Hypothesis 4: Participants in matched feedback conditions will perform better during training than participants in mismatched conditions*
>
> *Hypothesis 5: Participants in the hybrid conditions will perform better during training than participants in the ATI conditions*

Experimental Design Overview

In this study, participants learned a visuo-spatial periscope operation task called calling Angle on the Bow (AOB). This study employed a between subjects 2 (spatial ability; high spatial, low spatial) x 2 (feedback type; matched feedback, mismatched feedback) X 2 (adaptive approach; hybrid adaptive approach, ATI adaptive approach) factorial design. Participants were randomly assigned to one of the groups presented in Table 2 after completing a measure of spatial ability.

Table 2. Experimental conditions. The highlighted cells indicate conditions receiving more feedback.

Groups	Type of approach	Feedback type	Spatial ability
1	Hybrid	Match	High
2	Hybrid	Match	Low
3	Hybrid	Mismatch	High
4	Hybrid	Mismatch	Low
5	ATI	Match	High
6	ATI	Match	Low
7	ATI	Mismatch	High
8	ATI	Mismatch	Low

Independent Variables

In this experiment there were three independent variables, two (type of approach and feedback amount) were manipulated while the third, spatial ability, was a measured subject variable.

Adaptive Approach

Two adaptive approaches were be used in this experiment; participants either received an ATI approach or a hybrid approach. The approaches differ in what variable was used to provide feedback to trainees. In the ATI groups (groups 1, 2, 5, and 6), trainees either received the higher or less detailed feedback and the amount of feedback did not change throughout the training trials. In other words, participants in the less feedback ATI groups (groups 2 and 6) received less feedback throughout the training and participants in the more feedback ATI groups (groups 1 and 5) received more feedback throughout the training.

In the hybrid adaptive groups (groups 3, 4, 7, and 8), participants' initial feedback type was either less (groups 3 and 7) or more feedback (groups 4 and 8) similar to the ATI groups. However, after the first 45 periscope calls were made, feedback presentation was based on the participant's performance. More specifically, on calls 46-60, the feedback provided to trainees was based on the participant's average performance (on accuracy) on calls 1-45. On calls 61-75, the feedback given to the trainee was based on the participant's average performance on calls 46-60, and on calls 76-90, participants received feedback based on their average performance on calls 61-75. For the matched hybrid groups (groups 4 and 7) the feedback received was matched to their performance. In these conditions, if the participant performed well for periscope calls 1-45, they received less feedback during calls 46-60. Alternatively, if the participant performed poorly on these calls, they received more feedback during calls 46-60. For the mismatched hybrid groups (groups 3 and 8), the feedback received after call 45 was based on the opposite of their performance. For example, if one of the participants in this group performed well on calls 1-45, they received more feedback on calls 46-60, and if they performed poorly on calls 1-45 they received less feedback on calls 46-60. For an illustration of the hybrid approach in this experiment, see Figure 3.

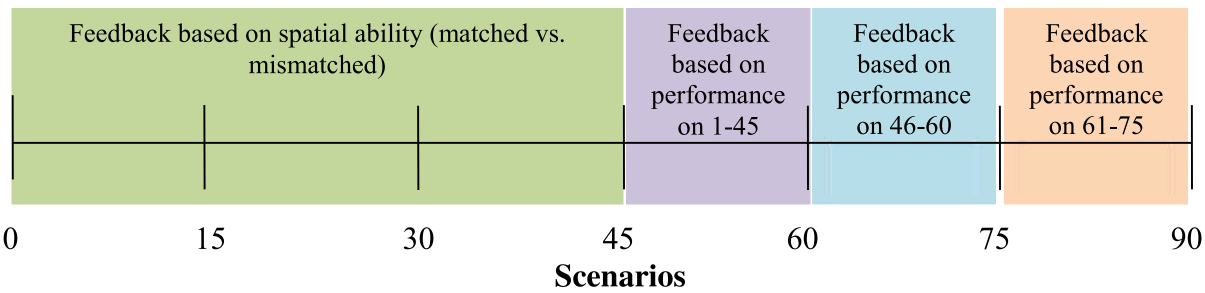


Figure 3: An illustration of the hybrid approach. Participants received feedback based on spatial ability for the first 45 scenarios, and feedback based on performance for scenarios 45-90.

Feedback

In this experiment, participants received either low or high amounts of feedback. More feedback was defined as process-type feedback given after a training scenario that provided the participant with tips to help them perform the AOB task better in future calls. For example, after a scenario that takes place at nighttime, a participant might get feedback that says "Notice the red light [on the ship]; this indicates that you are looking at the port side of the ship and the angle is between 0 and 112°". Participants who received process feedback were given two of these types of tips after each scenario. Accompanying the text feedback, there was also a visual representation of the call. The *call comparator* showed the trainee both a picture of what the ship looked like in the scenario and what the ship would have looked like if the participant's answer was correct (See Figure 4). Additionally, the feedback included in these conditions gave the participants numerical outcome feedback regarding their call. Above each picture in the call comparator, participants could see the call they made (over the picture on the left in Figure 4) and the correct answer (over the picture on the right in Figure 4).

The less feedback conditions received outcome feedback only; this was given after each scenario and provided the participant the correct answer as well as the answer they gave. The outcome feedback might have said, “You called port 135° when the correct answer is port 5°”. This feedback told the participant the correct answer, but did not give them any additional process information on how to improve their calls. I expected that participants with low spatial ability would benefit more from the process feedback tips and call comparator while participants with high spatial ability would perform better when they receive less feedback. While participants were measured on spatial ability prior to training, they were placed randomly in feedback conditions to create matched (groups 2, 4, 5, and 7) and mismatched (groups 1, 3, 6, and 8) conditions.

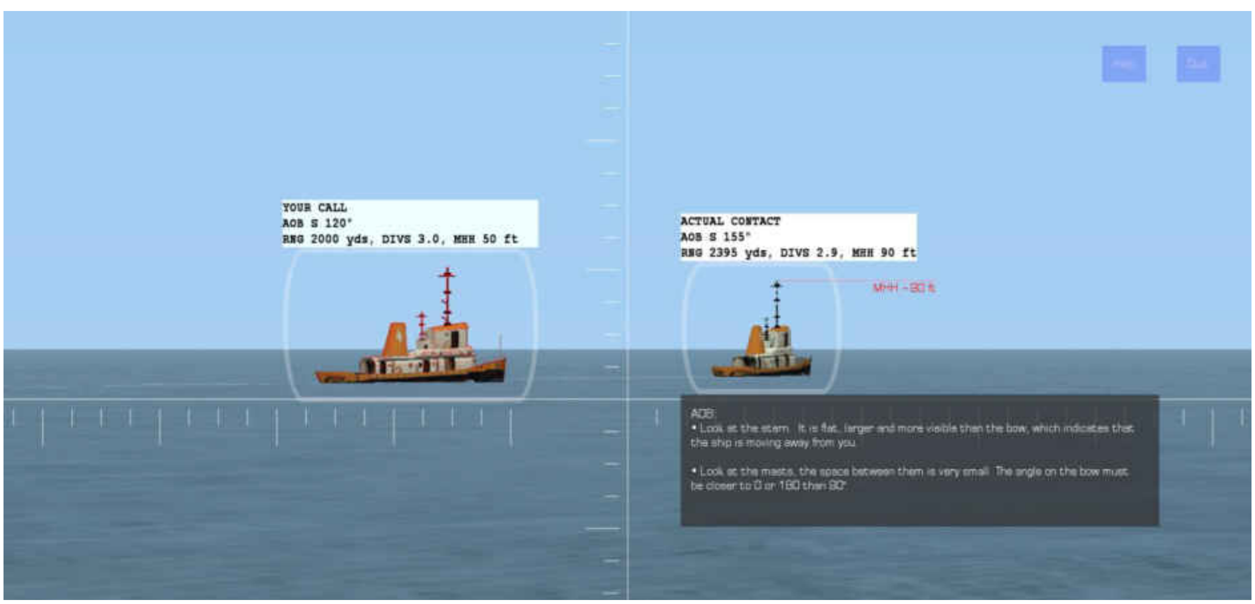


Figure 4: Screenshot of the call comparator and process tips

Spatial Ability

Spatial ability was a subject variable that was be measured by the Mental Rotation Test (Peters et al., 1995; Vandenberg & Kuse, 1978) described below. Cutoffs for high and low spatial ability were determined based on pilot data with a similar sample. Low spatial ability was defined as a score of 0-9 on the MRT and high spatial ability was as a score of 10-24 on the MRT. Participants were be assigned to conditions randomly, but matched on this variable to ensure equal numbers in each condition.

Method

Participants

A power analysis was performed to determine the number of participants required to reject the null hypothesis at an alpha level of .01. Power was set to .90, with an effect size of .41 based on the meta-analysis by Kluger and DeNisi (1996). Using the methods described by Cohen, Cohen, West and Aiken (2003), the power analysis yielded a result of 81 participants needed. In order to ensure each condition has the same number of participants and to increase the chances of finding a significant effect, sample size was increased 96, or 12 participants per group. Ninety-nine participants (43 males, 56 females, M_{age}=19.38, age range: 18-42 years) participated in the study.

Students were recruited through the participant collection software SONA Systems and were compensated with extra course credit for their participation. They were randomly assigned to one of the eight experimental groups (see Table 2). None of the participants had prior experience with the experimental task. All of the participants were treated according to guidelines set forth by the American Psychological Association (APA).

Four participants were removed from analysis because it was determined that they did not invest any effort in completing the scenarios during training (i.e. Christmas- tree-ing). If a participant typed in the same response for each scenario and/or their response time was under five seconds, they were removed from analysis. Of the participants removed for the above reason, one was in the Hybrid Matched High group, one was in the ATI Matched Low group, and two were in the Hybrid Matched Low group. Additionally, two outliers were removed from analysis after it was determined that they were two standard deviations below the mean for post-test median, gain, and post-test mean time. One of these participants was in the Hybrid Matched High group, and the other was in the Hybrid Matched Low group.

Tasks and Materials

Testbed and Apparatus

Testbed

The Periscope Operator Adaptive Trainer (POAT) was used as the experimental testbed. This testbed simulated a periscope operation task called Angle on the Bow. In the Angle on the Bow task, participants were asked to view different contacts (ships) and judge the angle that the contact is presenting in relation to their perspective on ownship. On a submarine, this information allows the periscope operator to determine if the contact being viewed is a collision threat to ownship. The testbed (See Figure 5) showed participants several scenarios, one at a time, each one containing one contact. The contact could be displayed to the participant at any angle between 0 and 180 degrees on the port (left) or starboard (right) side and it was the participant's job to determine the orientation of the contact. On the bottom of the screen, a box with several options allowed the participant to choose the side of the ship they are viewing (port, starboard, or N/A) and the angle of the contact.

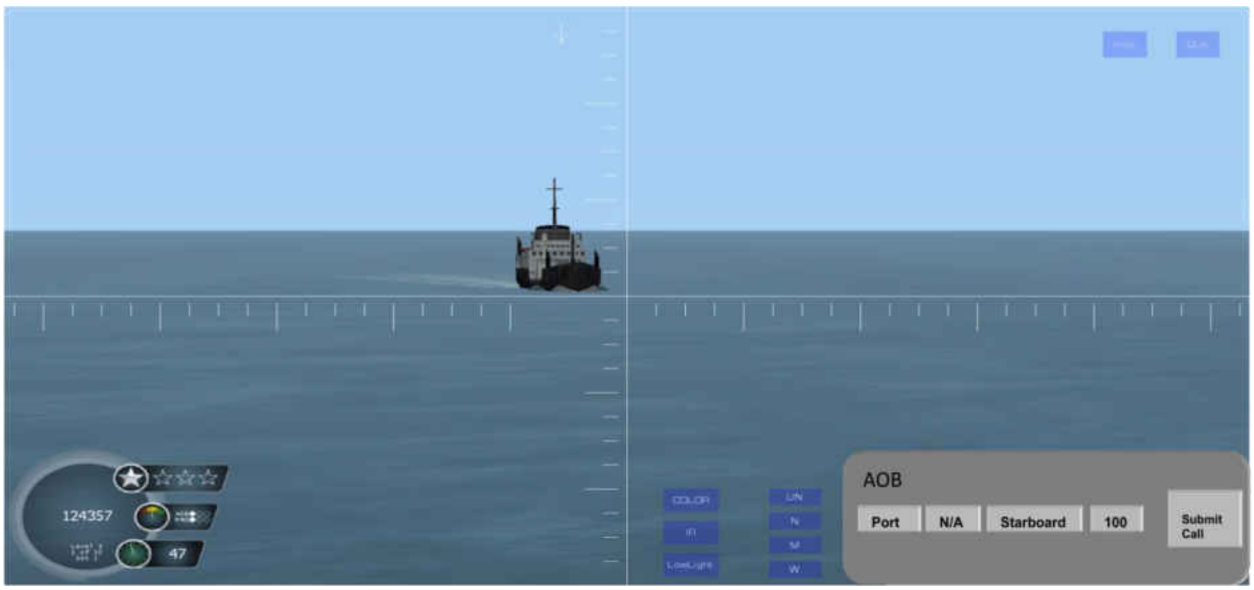


Figure 5: Screenshot of the experimental testbed interface

Participants selected the button representing the ship side they chose, and then entered the angle of the contact using the keyboard. Once they filled out the options in the box, the testbed gave the participants the option to either continue to the next scenario or take a break. Each scenario is displayed for one minute, after which it moves the participant on to the next scenario regardless of if they completed the task. The scenarios that the participants received varied between easy, medium, and hard. The difficulty of each scenario depended on a number of factors such as type of ship, angle of the ship, time of day (i.e., night, day, afternoon, etc.), sea state, and weather conditions. The difficulty of each scenario was determined empirically using pilot participants similar to the sample used in the present study. A recent study (Landsberg, Mercado, Van Buskirk, Lineberry, & Steinhauser, 2012) compared an adaptive version of this system to a non-adaptive version. The results of the study showed that participants who received the adaptive version of the training performed their calls significantly faster than the participants in the non-adaptive training group, and while not statistically significant, the average gain score from pre-test to post-test was higher for the adaptive group (31.94%) than the non-adaptive group (21.905%). The results of the abovementioned study indicate that this system can be used to train participants to more efficiently call angle on the bow when performance is used as the adaptation variable. In the present study, spatial ability was used to adapt the feedback that the trainees received from the system for participants in the ATI groups. The difficulty of scenarios presented to the participants was randomized within each set so that each participant saw the same scenarios in each set, but in random order. Each set contained easy, medium, and hard scenarios.

A near transfer task was also created in the POAT testbed. On the transfer task, participants were presented with different contacts than previously seen during training or in the pre/post-tests.

Equipment

POAT ran on two Dell Precision M6800 laptops, each containing an Intel® Core i5-4200M Processor, an AMD FirePro™ M6100 graphics card, and 8GB of DDr3L memory. Both laptops had 17.3" displays and ran the simulation at a resolution of 1600 x 900. Participants used a standard wired mouse and the on-board keyboard to complete the simulation.

Performance Measures

Performance measures in this experiment included the accuracy of the participant's call and the timing of the participants call. Accuracy in this experiment was defined as the absolute value of the delta between the actual angle of the target in the scenario and the angle called by the participant during the scenario. Delta was chosen rather than other measures (such as percent correct) because it gave a more precise representation of participants' performance. This is also the measure used by the Navy in their qualifications for periscope operators (e.g. prospective operators must make their AOB calls within 10° of the actual angle in order to qualify). Time to make the call was measured in milliseconds and started counting once the scenario begins and ended when the participant clicked on the Submit button. These scores were recorded for each call the participant made during the training session. Accuracy and timing were measured during the pre-test, training scenarios, post-test, and transfer test.

Manipulation checks

Measures of mental workload

The NASA TLX (Hart & Staveland, 1987) was used as one measure participants' subjective assessment of their mental workload on the task (see Appendix G). The NASA TLX is a subjective measure of workload that allows participants to rate their perceived workload based on several dimensions. The dimensions include mental demand, physical demand, temporal demand, performance, effort, and frustration level. In this experiment, the physical demand scale was removed because it was not relevant to task performance. Participants rated the remaining dimensions from low to high and were given a score between zero and one hundred (Gawron, 2000). The Mental demand scale of this measure served as an experimental manipulation check; trainees should have reported lower mental workload when they were in the conditions matched to their respective ability. Those with high spatial ability should report lower mental workload when they are in the condition with outcome feedback while those with lower spatial ability should report lower work load when they are in the condition with process feedback. The reverse should also be found; participants should rate their mental workload as higher when they are in conditions that are not matched to their ability.

In addition to the NASA TLX, the Paas (1992) 9-point mental effort rating scale was used to assess participants' subjective mental workload (see Appendix F). Unlike the NASA TLX, the 9-point rating scale is uni-dimensional and can be administered after each training task, rather than once at the end of a training session. Participants were asked to numerically rate their mental workload on a Likert scale ranging from 1: Very, very low mental effort, to 9: Very, very high mental effort after each phase of the training (i.e., once each after the practice session, training session, and the post-test). The 9-point mental effort rating scale was also used to calculate a score of instructional efficiency as discussed below.

Measure of instructional efficiency

When combined with performance data, results from the mental workload measures were used to create a measure of instructional efficiency. This measure is used by CLT researchers to compare different instructional strategies in terms of the amount of mental effort required to achieve a certain performance level (Paas & van Merriënboer, 1993). Taking both mental effort and performance into account gives a more precise description of the value of an instructional approach. For instance, if scores on the post-test of the matched and mismatched groups is equal, but their mental effort scores are significantly different, it indicates that the instruction was more efficient for the group with the lower mental effort scores. In this case, looking at the performance data alone would not have shown any value in choosing one type of instruction over the other.

Van Gog and Paas (2008) demonstrated that there are two types of measures of instructional efficiency- one that measures the efficiency of learning outcomes, and one that measures the efficiency of the learning process. The distinction between the aforementioned measures is that in the former, workload assessments pertain to the effort invested in achieving the results of the test itself whereas in the latter, mental workload assessments pertain to the effort invested to perform and complete learning tasks. In computational terms, the two types of instructional efficiency can be represented by the formulas below where P represents standardized test performance and E represents standardized scores of either test or learning related mental effort.

Instructional efficiency of outcomes:

$$Efficiency = \frac{zP_{test} - zE_{test}}{\sqrt{2}} \qquad (1)$$

Instructional efficiency of learning processes:

$$Efficiency = \frac{zP_{test} - zE_{learning}}{\sqrt{2}} \qquad (2)$$

While this measure is a sufficient indication of overall mental effort required to perform a task or complete a test, it did not differentiate between the three different types of cognitive load (germane, extraneous or intrinsic) espoused by CLT. Therefore, it cannot be used to interpret what kind of load the respondent was experiencing. As stated by van Gog and Paas (2008), "mental effort invested in the learning phase and mental effort invested in the test phase are very different." I expected trainees in all conditions to report relatively higher levels of mental effort during training, however, during the post-test I expected that participants who were in the matched groups would report lower mental effort scores. This is because mental effort ratings reported during the post-test should be reflective of the knowledge that the trainee has gained as a result of training and not the learning process.

Individual Difference Measures

Measure of spatial ability

A redrawn version of the Vandenberg and Kuse (1978) Mental Rotation Task (MRT) was used as the measure of mental rotation (Peters et al., 1995; see Appendix C). During the task, participants were asked to pick which two (out of a possible four) of the 3-dimensional geometric figures matched a comparator figure. Scores on the test could range from 1-24 as participants received one point for each question they answer correctly. An answer was only counted as correct if participants identified both of the matching figures; no partial credit was given. Based on previous testing using participants from the same population, a score of 0-9 was defined as low and a score from 10-24 was considered high spatial ability for this population.

Measure of working memory capacity

Shah and Miyake's (1996) spatial span task was used as a measure of working memory capacity (see Appendix D). In this task, participants were asked to mentally rotate a target figure to decide if it was presented normally or as a mirror image while also remembering the spatial orientation of each figure presented in the correct order. First, participants were presented with a letter that could appear at any orientation between 0° (upright) and 315°. The letter could also be presented as either normal or a mirror image. The participant was asked to say out loud and as quickly and accurately as possible if the letter was normal or a mirror image. The same letter was used throughout one trial, and each trial could consist of between two and five sets of letter presentations. At the end of a trial, a grid that represented the 8 possible orientations appeared on the screen and the participant was asked to recall the orientation of each letter in the set in the correct order. The task consisted of 20 letter sets with five sets at each level between two and five letters per set.

Measure of self-efficacy

A self-efficacy questionnaire was created for the purposes of this study (see Appendix I). It asked participants to rate on a five-point Likert scale from 1 (Not at all confident) to 5 (very confident) their confidence for performing tasks related to the training. I expected that trainees would report higher self-efficacy when they were in conditions that are matched to their ability. Likewise, I expected that participants in conditions not matched to their ability would report lower self-efficacy.

Procedure

In the beginning of each experimental session, participants were asked to read the informed consent. Following this, the experimenter gave a brief description of the schedule for the session and asked the participant to fill out questionnaires including a demographic questionnaire (see Appendix A) and the mental rotation test. At this time, the participants were administered the working memory capacity measure via PowerPoint. Participants were then randomly assigned to feedback conditions based on their spatial ability. The participant's spatial ability score was calculated from the mental rotation test and the participant was randomly assigned to one of the eight conditions. For example, if a participant's measured spatial ability was high and they were randomly assigned to the matched hybrid condition, they were placed in the hybrid condition where outcome feedback was presented for the first 45 scenarios. After they filled out the initial questionnaires, participants viewed a brief tutorial that described the Angle on the Bow task and then took a short quiz (see Appendix E) to ensure that they understood the material. Participants then began calling AOB using the POAT simulation. They completed a pre-test containing 30 scenarios, the experimental portion containing 90 scenarios, followed by a post-test with 30 scenarios that were the same as the scenarios presented in the pre-test but in a randomized order, and a transfer task containing 23 scenarios. After each portion of the training, participants filled out the 9-point mental workload questionnaire. Following the experimental portion and the post-test participants filled out the NASA TLX. After the transfer portion of the experimental task, participants were asked to fill out the self-efficacy questionnaire. At the end of the experiment participants were debriefed. See Table 3 for an overview of the experimental procedure.

Table 3. Overview of Experimental Procedure

Activity	Time (minutes)
Informed consent and pre-brief	5
Demographics questionnaire	5
Mental Rotation Test	10
Working memory capacity test	10
Task/testbed familiarization	15
Knowledge quiz	10
Pre-test	15
9-point mental effort questionnaire	2
Training scenarios	30
9-point mental effort questionnaire	2
NASA-TLX	5
Optional break	5
Post-test	15
9-point mental effort questionnaire	2
NASA-TLX	5
Transfer	10
Self-Efficacy Questionnaire	5
Total Time	151

CHAPTER THREE: RESULTS

Manipulation checks

Manipulation checks were used to examine if random assignment to groups was achieved and if the manipulations of the independent variable functioned as expected. Two manipulation checks were performed to determine whether the experimental groups were equal prior to the experimental manipulation. Firstly, in order to check that assignment to groups was random, an Analysis of Variance (ANOVA) was used to examine if the eight experimental groups were similar demographically. There were no significant differences between the groups on the demographic variables: Age: $F(7, 84)= 1.85$, $p=.09$; Gender: $F(7, 84)= .954$, $p=.47$; Handedness: $F(7,84)= .662$, $p=.70$; Frequency of PC use $F(7,84)= .752$, $p=.63$; Experience with computers $F(7,84)=1.4$, $p=.20$; Hours per week playing video games $F(7,84)=.672$, $p=.69$; Experience with first-person perspective video games $F(7,83)=1.1$, $p=.35$, Experience with third-person perspective video games: $F(7,83)=.95$, $p=.47$; Experience with solving picture puzzles: $F(7, 84)= 1.55$, $p=.16$, and Experience with sculpture, painting, drawing or other visual arts $F(7,83)=1.5$, $p=.18$. Table 4 shows the means and standard deviations for each demographic variable.

Table 4. Means and standard deviations for demographic variables

	1	2	3	4	5	6	7	8
Age	20.36	18.20	19.00	22.58	18.18	18.69	20.00	18.75
	(6.12)	(.42)	(1.41)	(6.70)	(.40)	(1.65)	(3.97)	(1.76)
Gender	1.45	1.90	1.58	1.58	1.45	1.53	1.45	1.67
	(.52)	(.32)	(.51)	(.51)	(.52)	(.52)	(.52)	(.49)
Handedness	2.00	2.00	1.83	1.83	1.91	2.00	1.83	1.83
	(00)	(00)	(.39)	(.39)	(.30)	(00)	(.39)	(.39)
Frequency	6.18	6.00	6.33	6.25	6.55	6.15	6.27	6.42
of PC use	(.60)	(.67)	(.49)	(.62)	(.52)	(.69)	(.65)	(.79)
Computer	2.90	2.50	2.50	2.75	2.91	2.46	2.90	2.83
experience	(.70)	(.71)	(.52)	(.45)	(.70)	(.52)	(.54)	(.39)
Video game	4.09	.90	4.16	1.83	4.72	1.92	3.72	3.50
hours/week	(10.64)	(1.29)	(5.82)	(3.10)	(4.63)	(4.27)	(3.71)	(6.02)
1st-person	3.18	2.20	2.83	2.25	3.54	2.46	2.9	2.67
games	(1.66)	(1.75)	(1.40)	(1.28)	(1.37)	(1.45)	(1.20)	(1.37)
3rd- person	2.90	3.40	3.00	2.5	3.81	3.00	3.30	3.00
games	(1.70)	(1.43)	(1.27)	(1.24)	(1.33)	(1.08)	(1.25)	(1.41)
Picture	3.18	4.00	3.42	2.92	3.27	3.15	3.20	3.08
puzzles	(.87)	(.94)	(.51)	(.67)	(1.10)	(.99)	(.92)	(.67)
Sculpture,	2.63	2.90	1.92	2.34	2.82	2.08	2.00	2.92
painting,	(1.36)	(1.20)	(.90)	(1.15)	(1.17)	(1.32)	(.94)	(1.08)
visual arts								

Note. Gender is dummy coded where 1=male, 2=female. Handedness 1=left, 2=right. Frequency PC use: 1= I've never worked with a PC, 2= Only a couple of times ever, 3=Several times a year, 4=Several times a month, 5= Several times a week, 6=At least once a day, every day, 7=For Several hours every day . Computer experience: 1 = No experience, 2 = Know a little (internet, Microsoft programs), 3 = Know quite a bit (e.g., other software, some programming), 4 = Expert (e.g., multiple software packages, multiple programming languages). 1st-person game experience, 3rd-person game experience, picture puzzle experience, sculpture, painting, visual arts: 1 = Not at all experienced, 2 = Somewhat experienced, 3 = Very experienced.

Secondly, participants' delta scores on the pre-test were used as a measure to ensure that the experimental groups were randomly assigned. Medians for pre-test scores were used because participant deltas for each scenario could range from 0 to 180° and median is less susceptible to extreme values. Prior to the pre-test, none of the participants reported having any experience on the experimental task, and no feedback was provided during the pre-test. A difference was expected between participants of higher and lower ability and indeed an ANOVA that included all eight groups revealed that there were significant differences between the groups on pre-test $F(7,84)=3.16$, $p=.005$, $\eta^2=.21$. LSD post hoc tests revealed that the ATI Mismatched High ($M=28.27$, $SD=8.13$) group performed significantly better than the Hybrid Matched Low group ($M=44.90$, $SD=9.34$) and the ATI matched Low group ($M=44.80$, $SD=9.21$) [Note: lower means indicate better performance]. A t-test that compared higher and lower ability participants regardless of approach and feedback manipulation revealed a significant difference between the high and low spatial ability group, $t(90)=2.78$, $p=.007$, with the higher ability group ($M=34.56$, $SD=13.93$) performing better than the lower ability group ($M=41.22$, $SD=9.26$). This represented a medium effect, $d=-.59$. However, no significant differences were present between the matched and mismatched groups and the ATI and hybrid groups when entered in a t-test, $t(90)=-1.72$, $p=.09$, $d=-.36$, $t(90)=.52$, $p=.61$, $d=.11$ respectively. Additionally, there were no significant differences on pre-test time, $F(7,84)=.56$, $p=.79$, $\eta^2=.04$. The means and standard deviations for pre-test deltas and times can be seen in Table 5. Because there were pre-test differences between participants of higher and lower spatial ability, pre-test scores were used as a covariate in subsequent analyses.

Table 5. Means and standard deviations for pre-test median deltas and pre-test time

Condition	Pre-test median deltas*	Pre-test time
Hybrid match high	31.55 (5.31)	20.71 (3.77)
Hybrid Match low	44.90 (9.34)	18.13 (3.61)
Hybrid mismatch high	34.92 (8.87)	19.11 (6.94)
Hybrid mismatch low	37.92 (8.92)	17.98 (5.65)
ATI Matched High	41.86 (23.93)	21.21 (10.77)
ATI matched low	44.80 (9.21)	18.14 (6.44)
ATI mismatch high	28.27 (8.13)	21.38(7.26)
ATI mismatch low	37.71 (7.39)	18.76 (4.46)

**Note: Lower numbers indicate better performance*

Next, checks were performed to ensure the manipulations of the independent variables behaved as intended. As mentioned above, there was a significant difference between the median pre-test scores of the higher and lower ability participants. To ensure the cutoffs for high and low spatial ability determined prior to experimentation were accurate for the current sample, the median for scores on the Mental Rotation Test was calculated. The median score on the MRT was 9 (SD=5.10) indicating the cutoff was reasonable for this sample.

A visual comparison of participants' path through training was performed. This was done to verify that the independent variable Approach worked as intended. In the ATI conditions, participants should have received the same type of feedback (based on their spatial ability) throughout training. In the hybrid conditions, participants should have received the same feedback (based on their spatial ability) for the first 45 scenarios, and then different feedback for the 45^{th}-90^{th} scenarios based on their performance. Looking at the path participants took through training showed that the ATI conditions and the hybrid conditions for higher ability participants worked as expected. However, the lower ability hybrid groups did not get multiple types of feedback as was expected: the Hybrid Matched Low group received only the more detailed feedback with the exception of two participants who received the lower amount for one set of scenarios each, both in scenarios 60-75. Because of this, this condition was equal to the ATI Matched Low group. Similarly, the Hybrid Mismatch Low group received only the less detailed feedback, with the exception of two participants who received the more detailed feedback for one set of scenarios each (also on scenarios 60-75). This group, then, was almost identical to the ATI Mismatch Low group. Therefore no comparisons will be performed between the low ability ATI and Hybrid groups. This impacted hypotheses 2b and 5 because the lower ability Hybrid and ATI groups could not be compared. Table 6 shows the comparison between the different conditions and the type of feedback they received during training.

Table 6. Comparison between conditions based on feedback

Condition	Scenarios 1-45	Scenarios 45-60
Hybrid Match High	Less	Changed based on performance
ATI Matched High	Less	Less
Hybrid Mismatch High	More	Changed based on performance
ATI Mismatch High	More	More
Hybrid Match Low	More	More
ATI Matched Low	More	More
ATI Mismatch Low	Less	Less
Hybrid Mismatch Low	Less	Less

Scores on the mental effort rating scale and NASA TLX mental effort scale were used to determine if participants reported similar mental effort ratings at pre-test and if providing participants with feedback matched to their ability reduced subjective metal effort during the post-test. As mentioned previously, the 9-point mental effort scale was administered three times: 1. After the pre-test, 2. After training, and 3. After the post-test. The means and standard deviations for participant mental effort ratings can be seen in Table 7. There were no significant differences between groups on mental effort questionnaires at time 1, $F(7,84)=1.44$, $p=.20$, $\eta^2=.11$, suggesting that participants were similar at pre-test, and there were also no significant differences between the groups at time 3, $F(7,84)=1.27$, $p=.27$, $\eta^2=.10$. The latter result was not expected; theoretically, the matched groups should have reported lower mental effort on the post-test, having created better schemas during the training session that would help them reduce the amount of cognitive resources expended at this point. This result may indicate that the feedback manipulation was not effective in reducing extraneous cognitive load during training. Further, there was a significant difference between the groups on the second mental effort questionnaire that was administered after the training session $F(7,84)=2.16$, $p=.046$, $\eta^2=.15$ (See Figure 6). Post hoc tests (LSD) revealed that the Hybrid Matched Low group rated their mental effort on the post-test significantly higher than the Hybrid Mismatched High, the ATI Mismatched High, and the ATI Mismatched Low groups. There was also a significant difference between the Hybrid Mismatched High group and the ATI Matched High group with the ATI Matched High group rating their mental effort as higher. Lastly, the Hybrid Matched Low group rated their mental effort significantly lower than the Hybrid Mismatched Low group. Figure 7 shows the mental effort scores for time 2 mental effort ratings. That several of the matched groups rated their mental effort as significantly higher than mismatched groups indicates that the manipulation of feedback match/mismatch may not have had the intended effect of decreasing cognitive load. However, there is no way to break down what type of load is being measured (i.e. extraneous, intrinsic, or germane), and therefore it is possible that participants in matched conditions were experiencing higher levels of germane load during training. This idea was explored further in the analyses of efficiency of outcomes and learning below.

Table 7. Means and standard deviations for the subjective mental effort ratings

Condition	1. After pre-test	2. After training	3. After post-test
Hybrid match high	6.00 (1.55)	6.27 (1.10)	6.36 (1.63)
Hybrid Match low	6.50 (.85)	7.40 (1.07)	6.40 (1.26)
Hybrid mismatch high	6.25 (1.54)	6.00 (1.41)	6.00 (1.28)
Hybrid mismatch low	5.54 (1.72)	6.25 (1.96)	5.08 (1.88)
ATI Matched High	6.82 (.75)	7.27 (.90)	6.36 (1.57)
ATI matched low	5.92 (1.19)	6.77 (1.23)	5.77 (.83)
ATI mismatch high	5.73 (1.00)	5.91 (1.37)	5.64 (1.29)
ATI mismatch low	5.50 (1.31)	5.83 (1.58)	5.67 (1.07)

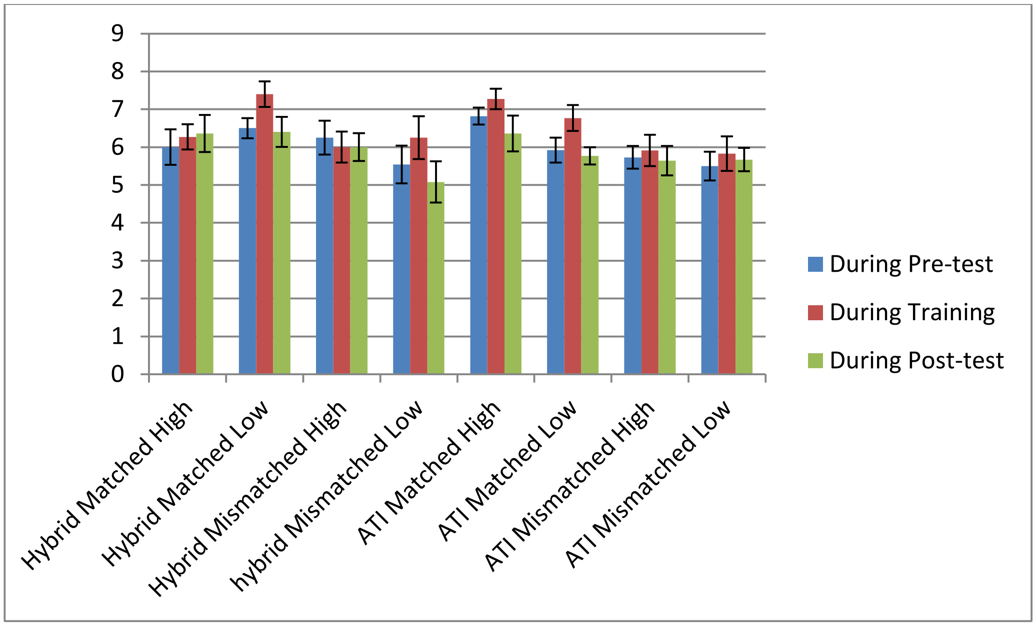


Figure 6: Mean mental effort ratings of the eight experimental groups

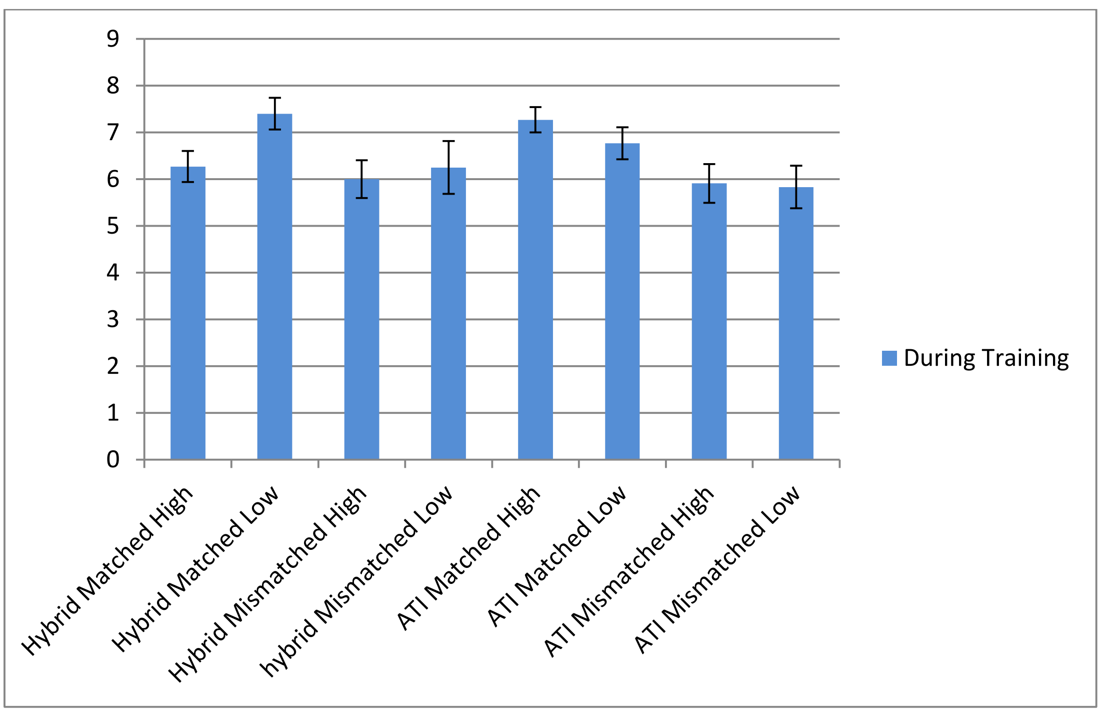


Figure 7: Mean mental effort ratings of the eight experimental groups for the second administration

Participants' scores on the Mental Demand scale of the NASA TLX were also examined to determine if there were differences between the groups. The NASA TLX was administered twice: 1. After the training session, and 2. After the post-test. Table 8 shows the means and standard deviations for the groups' answers on the Mental Demand scale at times 1 and 2. A one-way ANOVA revealed that there was a significant difference between the groups on their scores at the first TLX administration, $F(7,84)=2.28$, $p=.035$, $\eta^2=.16$. Post-hoc tests (LSD) revealed that the ATI Matched High group rated their mental demand significantly higher than Hybrid Matched High group. Additionally, the Hybrid Matched Low group rated their mental demand significantly higher than the Hybrid Mismatched High, Hybrid Mismatched Low, ATI Mismatched High and the ATI Mismatched Low groups. The Hybrid Mismatched High group rated their mental demand significantly lower than the Hybrid Matched Low and the ATI Matched High group. The Hybrid Mismatched Low group rated their mental effort significantly lower than the ATI Matched High group. The ATI Matched High group rated their mental demand significantly higher than the ATI Mismatched High and ATI Mismatched Low groups. There were no significant differences between the groups at time 2, $F(7,84)=.43$, $p=.879$, $\eta^2=.03$. Figure 8 shows the representation of the groups' scores on the NASA TLX Mental Demand Scale. As was seen in the results of the 9-point mental effort rating at the time of training, several of the mismatched groups rated their mental effort as lower than participants in the matched groups. This adds more evidence that the CLT manipulation may not have worked as hypothesized in decreasing extraneous load.

Table 8. Means and standard deviations on NASA TLX times 1 and 2

Condition	1	2	3	4	5	6	7	8
Mental Demand Time 1	64.91 (9.98)	78.70 (12.29)	60.83 (19.03)	59.67 (23.18)	79.82 (12.26)	67.69 (17.46)	60.82 (17.43)	49.27 (23.29)
Mental Demand Time 2	57.55 (18.17)	60.50 (26.77)	56.83 (21.66)	54.58 (27.50)	64.55 (22.59)	53.23 (26.09)	62.42 (22.86)	56.25 (18.16)

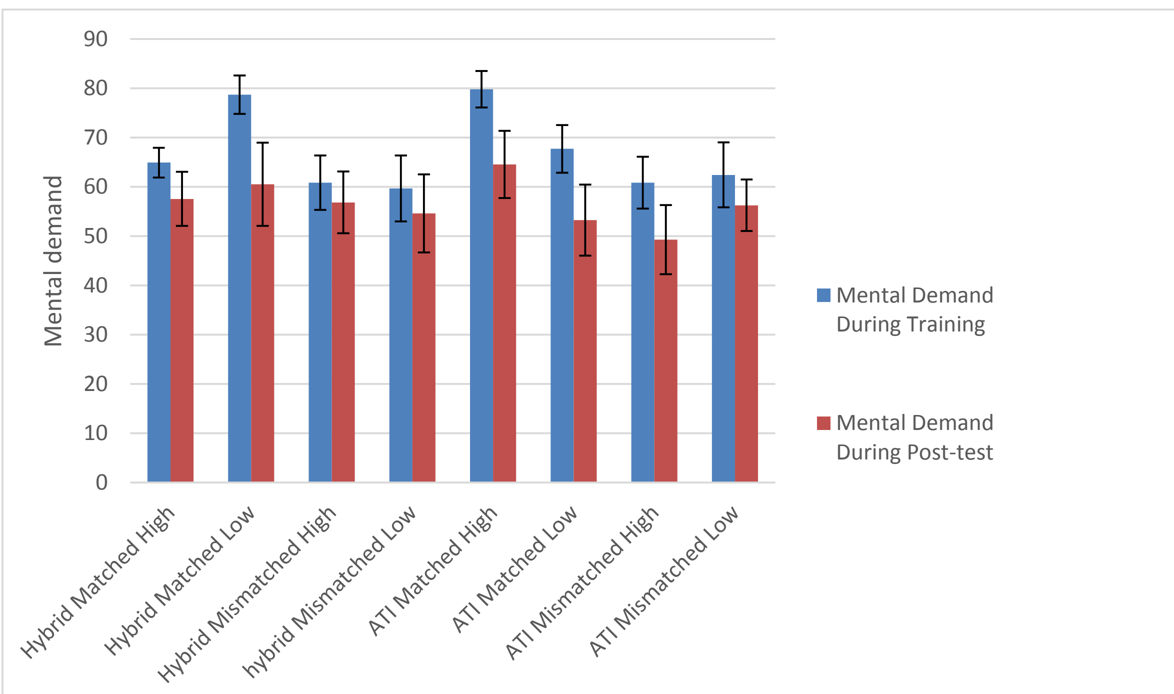


Figure 8: Subjective ratings of Mental Demand on the NASA TLX

As mentioned previously, an instructional efficiency score was also calculated for each group, both for the efficiency of the actual learning process and the efficiency of outcomes (See Equations 1 and 2). For efficiency of the learning process, participants' scores on post-test and their mental effort ratings after the training phase were standardized and placed into Equation 2. For efficiency of outcomes, participants' ratings on mental effort questionnaire that were filled out after the post-test were subtracted from their standardized post-test deltas, shown in Equation 1. The means and standard deviations of participants' efficiency scores for both outcomes and the learning process can be seen in Table 9. There were no significant differences between groups on either Efficiency of Outcomes, $F(7, 84)=2.03$, $p=.06$, partial $\eta^2=.14$, or Efficiency of Learning Processes, $F(7, 84)=.79$, $p=.60$, partial $\eta^2=.06$. These results do not support the existence of an efficiency of the Matched conditions, or the Hybrid conditions as was expected. In fact, when examining the means, it appears that the Hybrid Matched High group had the lowest score for Efficiency of outcomes. This indicates that the manipulation of Match/Mismatched feedback did not work as expected.

Table 9. Means and standard deviations of participant efficiency scores

Condition	1	2	3	4	5	6	7	8
Efficiency of outcome	-.61 (.78)	-.07 (.77)	-.32 (.67)	.58 (1.38)	-.12 (1.32)	.57 (1.26)	-.23 (.68)	.06 (.81)
Efficiency of learning	-.28 (.57)	-.28 (.67)	-.05 (.69)	.27 (1.57)	-.29 (.99)	.35 (1.45)	-.10 (.70)	.25 (1.01)

Hypothesis testing

Hypothesis 1: Participants in matched feedback conditions will perform better than those in mismatched feedback conditions

> *Prediction a: High spatial ability trainees who are given less feedback (matched) will perform better than high spatial ability trainees who are given more feedback (mismatched)* (Groups 1, 5 > 3, 7)

In order to test the hypothesis that high ability trainees who were given matched feedback would perform better than high ability trainees who were given mismatched feedback, an Analysis of Covariance (ANCOVA) was performed for each dependent variable. The groups were collapsed on the variable of feedback type creating a High Spatial Matched Group and a High Spatial Mismatched Group. Table 10 shows the means for the high spatial ability groups on post-test median, post-test time, transfer median, transfer time, pre- to post-test gain scores, instructional efficiency of outcomes, and instructional efficiency of the learning process. The Select Cases function in SPSS was used to isolate the higher ability participants only for these analyses. Although the analyses included only two groups, ANCOVA was chosen so that the covariates for each analysis could be accounted for. The first analysis was performed on post-test median. The covariate (pre-test median) was significantly related to post-test median $F(1, 42) = 44.91$, $p < .0001$, partial $\eta^2 = .52$. There was no significant effect of match/mismatch on post-test median, $F(1, 42) = .115$, $p = .74$, partial $\eta^2 = .003$.

For post-test mean time an ANCOVA was performed using pre-test time as a covariate. The covariate was significantly related to post-test time, $F(1,43) = 28.71$, $p < .0001$, partial $\eta^2 = .40$. There were no significant differences between the groups on post-test time, $F(1,43) = .04$, $p = .85$, partial $\eta^2 = .001$.

For transfer median, ANCOVA was performed using pre-test median as a covariate. The covariate (post-test median) was significantly related to transfer median, $F(1,43)=42.44$, $p<.0001$, partial $\eta^2=.20$. There was no significant effect of match/mismatch on transfer median $F(1,43)=.059$, $p=.81$, partial $\eta^2=.001$.

In order to find differences in transfer time mean an ANCOVA was performed using post-test time as a covariate. The covariate was significantly related to the DV, $F(1,43)= 51.22$, $p<.0001$, partial $\eta^2=.52$. Match/mismatch was not significantly related to transfer time, $F(1,43)=.16$, $p=.69$, partial $\eta^2=.006$.

Next, a t-test was performed to determine if there was a difference between the groups on pre- to post-test median gain. Gain scores were calculated by subtracting both the pre-test and post-test from 180° (the maximum possible delta), and then using the following equation to calculate the gain score:

Pre to post gain score:

$$Pre/Post\ Gain = \frac{post-pre}{180-pre} \tag{3}$$

There was no significant difference in the gain scores between high ability participants who received matched feedback and those that received mismatched feedback on pre to post-test gain, $t(44)=-.14$, $p=.91$, $d=.09$.

Lastly, an analysis was also performed to ascertain if there was an instructional efficiency of outcomes and of learning outcomes for these groups. Their instructional efficiency (see above for calculation of these scores) scores were analyzed using an ANOVA with efficiency scores as the dependent variable. For efficiency of outcomes, there were no significant differences between the groups, $F(2,42)= .61$, $p=.44$, partial $\eta^2=.014$. Similarly for efficiency of learning processes, there were no significant differences between the groups, $F(2,42)= 2.80$, $p=.10$, partial $\eta^2=.06$.

Hypothesis 1 prediction A was not supported.

Table 10. Means and standard deviations for the high ability participants

Condition	Post-test median*	Post-test mean time (seconds)	Transfer median*	Transfer mean time (seconds)	Pre to post gain scores	IE outcome	IE learning
High Spatial	22.82	12.82	28.95	14.04	.26	-.36	-.28
Matched	(16.22)	(4.01)	(16.27)	(8.62)	(.19)	(1.09)	(.79)
High Spatial	18.85	12.67	23.92	13.13	.32	-.28	-.07
Mismatched	(5.95)	(3.36)	(17.34)	(3.30)	(.21)	(.66)	(.68)

**Note: Lower numbers indicate better performance.*

> *Hypothesis 1 Prediction b: Low spatial ability trainees who are given more feedback (matched) will perform better than low spatial ability trainees who are given less feedback (mismatched)* (Groups 2, 6 > 4, 8)

Similarly to prediction a, ANCOVAs were used to test the hypothesis that lower ability trainees who received matched feedback would perform better than lower ability trainees who received mismatched training. The Select Cases function in SPSS was used to isolate the lower ability participants only for these analyses. Groups were collapsed on the variable of feedback Match/Mismatch creating a Low Spatial Matched Group and a Low Spatial Mismatched Group. Table 11 shows the means and standard deviations for the low spatial ability groups on post-test median, post-test time, transfer median, transfer time, pre- to post-test gain scores, efficiency of outcomes, and efficiency of the learning process. The analysis for post-test median was performed first. The covariate (pre-test median) was significantly related to post-test median, $F(1,44)=5.26$, $p=.03$, partial $\eta^2=.11$, however there was no significant effect of match/mismatched feedback on post-test median, $F(1,44)=.20$, $p=.66$, partial $\eta^2=.004$.

Next, an ANCOVA was performed on post-test timing. The covariate (pre-test time mean) was significantly related to the DV, $F(1,43)=17.94$, $p<.0001$, partial $\eta^2=.20$. Match/mismatch was not significantly related to post-test time, $F(1,43)=.53$, $p=.47$, partial $\eta^2=.008$.

Another ANCOVA was used to test the means for transfer median. The covariate (pre-test median) was significantly related to the DV, $F(1,42)=24.09$, $p<.0001$, partial $\eta^2=.14$. Match/mismatch was not significantly related to transfer median, $F(1,42)=1.34$, $p=.26$, partial $\eta^2=.019$.

In order to test the hypothesis for transfer timing, an ANOVA was performed using post-test time as the covariate. Post-test time was significantly related to the DV, $F(1,42)=37.52$, $p<.0001$, partial $\eta^2=.22$. Match/mismatch was not significantly related to transfer time, $F(1,42)=.34$, $p=.56$, partial $\eta^2=.002$.

For the pre-to post-test gain differences a t-test was used to determine if the matched low spatial group performed better than the mismatched low spatial group on gain scores. There was no significant difference between the groups, $t(43)=-.12$, $p=.90$, $d=.15$.

Analyses were also performed to examine instructional efficiency between the groups. The efficiency of outcomes was analyzed first. The scores for the two groups were entered into an ANOVA with efficiency of outcomes as the dependent variable. There were no significant differences between the groups on this variable, $F(2,44)= 1.74$, $p=.19$, partial $\eta^2=.038$. Similar analyses were performed for efficiency of the learning process, and again it was found that there were no differences between the groups, $F(2,44)= .53$, $p=.47$, partial $\eta^2=.012$.

Hypothesis 1b was not supported.

Table 11. Means and standard deviations for lower ability participants

Condition	Post-test median*	Post-test mean time (seconds)	Transfer median*	Transfer mean time (seconds)	Pre to post gain scores	IE outcome	IE learning
Low Spatial	33.04	11.54	37.08	11.80	.18	.29	.07
Matched	(17.40)	(4.00)	(21.66)	(3.92)	(.23)	(1.10)	(1.19)
Low Spatial	26.72	12.25	28.34	11.67	.22	.33	.26
Mismatched	(13.75)	(4.02)	(12.67)	(2.70)	(.18)	(1.14)	(1.29)

**Note: Lower numbers indicate better performance*

Hypothesis 2: Participants who receive hybrid adaptive training will perform better than those who receive ATI adaptive training

> *Prediction a: High ability participants who receive hybrid training will perform better than high ability participants who receive ATI training* (Groups 1, 3 > 5, 7)

In order to test the hypothesis that high ability participants who received hybrid training would perform better than those who received ATI training, several ANCOVAs were performed. The means and standard deviations for these groups can be found in Table 12. For prediction A, the Select Cases function in SPSS was used to isolate only the higher ability participants and the Groups were collapsed on the variable of AT Approach. This created a High Spatial Hybrid group and a High Spatial ATI Group. The first ANCOVA was performed using post-test median as the dependent variable. The analysis revealed that there were no post-test differences between the groups, $F(1,43)=1.18$, $p=.28$, partial $\eta^2=.03$. The covariate, pre-test performance was significantly related to post-test means, $F(1,43)=47.28$, $p<.0001$, partial $\eta^2=.53$.

An ANCOVA with pre-test time mean as a covariate revealed that pre-test mean time was significantly related to post-test mean time, $F(1,43)=31.41$, $p<.0001$, partial $\eta^2=.49$. There was also a significant difference between the groups on post-test time, $F(1,43)=4.12$, $p=.049$, partial $\eta^2=.09$, where the ATI group performed their calls on the post-test significantly faster (M=12.17, SD=4.11) than the Hybrid group, (M=13.30, SD=3.14).

For transfer median, an ANCOVA with post-test as the covariate revealed that pre-test scores were significantly related to transfer median scores, $F(1,43)=49.60$, $p<.0001$, partial $\eta^2=.54$. There were no significant differences between the groups on transfer median scores $F(1,43)=2.17$, $p=.148$, partial $\eta^2=.05$.

In order to examine differences on transfer time, an ANCOVA was used with post-test time as the covariate. The covariate was significantly related to the DV, $F(1,43)=53.86$, $p<.0001$, partial $\eta^2=.56$. Approach was not significantly related to transfer time, $F(1,43)=.69$, $p=.41$, partial $\eta^2=016$.

A t-test was performed to compare the gain scores of high ability participants who received hybrid training to those that received ATI training. There was no significant difference between the groups, $t(44)=-1.06$, $p=.30$, $d=.32$.

Learning efficiency scores were also calculated using an ANOVA. There were no significant differences between the groups on instructional efficiency of outcomes, $F(2,43)= 1.15$, $p=.29$, partial $\eta^2=.026$, or on instructional efficiency of the learning process, $F(2,43)= .02$, $p=.89$, partial $\eta^2=.001$

Hypothesis 2a was not supported.

Table 12. Means and standard deviations for the high ability Hybrid and ATI groups

Condition	Post-test median*	Post-test time (seconds)	Transfer median*	Transfer time (seconds)	Pre to post gain scores	IE outcomes	IE learning
High Spatial	18.80	13.30	26.87	13.77	.42	-.46	-.16
Hybrid	(5.46)	(3.14)	(17.70)	(4.13)	(.18)	(.72)	(.63)
High Spatial	22.78	12.17	25.86	13.36	.35	-.18	-.19
ATI	(16.01)	(4.11)	(16.70)	(8.07)	(.22)	(1.02)	(.84)

**Note: Lower numbers indicate better performance*

*Prediction b: Lower ability participants who receive hybrid training will perform better than lower ability participants who receive ATI training (*Groups 2, 4 > 6, 8)

Prediction B could not be tested, because as mentioned above, the lower ability hybrid group essentially received the same training as lower ability participants who received ATI training. Therefore, this analysis was excluded.

Prediction c: Participants who receive hybrid matched training will do better than all other groups (Group 1 > 3, 5, 6, 7, 8)

Prediction C hypothesized that the Hybrid matched group would perform better than other groups. Groups 2 and 4 were not included in this analysis because the lower ability Hybrid groups were very similar to the lower ability ATI groups. The remaining groups' scores on post-test median, post-test time, transfer median, transfer time, pre to post- gain scores, instructional efficiency of outcomes, and instructional efficiency of the learning process can be found in Table 13. The Select Cases function in SPSS was used to isolate the group mentioned above and ANCOVAs were used to test if the groups showed differences.

The first ANCOVA was performed to examine the differences between the groups on post-test median scores. Pre-test median was entered as the covariate, and it was found to be significantly related to scores on the post-test, $F(1, 63)=14.91$, $p<.0001$, partial $\eta^2=.19$. Condition was not related to post-test performance, $F(5, 63)=.1.26$, $p=.29$, partial $\eta^2=.09$.

In the next ANCOVA, post-test mean time was entered as the dependent variable. Pre-test time was used as the covariate, and it was significant, $F(1, 63)=37.99$, $p<.0001$, partial $\eta^2=.38$. Condition was not related to post-test mean time, $F(5, 63)=.72$, $p=.61$, partial $\eta^2=.05$.

Next, transfer median was examined. The covariate, pre-test performance, was significantly related to the dependent variable, $F(1, 63)=8.00$, $p =.006$, partial $\eta^2=.11$. Condition was not related to transfer median scores, $F(5, 63)=.39$, $p=.85$, partial $\eta^2=.03$.

Another ANCOVA was performed for transfer-time using pre-test time as the covariate. Pre-test times were significantly related to transfer time performance, $F(1, 63)=72.96$, $p<.0001$, partial $\eta^2=.53$, however condition was not, $F(5, 63)=.65$, $p=.66$, partial $\eta^2=.05$.

An ANOVA was used to examine differences between the groups on their pre to post-test gain scores. There was no significant difference between the groups, $F(5, 64)=.1.22$, $p=.31$, partial $\eta^2=.09$.

Lastly, analyses were performed in order to examine the relative instructional efficiencies for each group. The groups' means on instructional efficiency of outcomes and efficiency of learning were entered into a one-way ANOVA. There were no significant differences between the groups for efficiency of outcomes, $F(5, 64)=.2.10$, $p=.08$, $\eta^2=.16$. There was also no significant difference between the groups for efficiency of the learning process, $F(5, 64)=.91$, $p=.48$, $\eta^2=.07$.

Table 13. Means and standard deviations for all groups (except the low ability hybrid groups)

Condition	Post-test median*	Post-test time (seconds)	Transfer median*	Transfer time (seconds)	Pre to post gain scores	IE outcomes	IE learning
1	17.73	13.60	25.91	14.11	.42	-.61	-.28
	(6.01)	(3.12)	(8.96)	(5.12)	(.14)	(.78)	(.57)
3	19.79	13.01	27.76	13.46	.42	-.32	-.05
	(4.96)	(3.27)	(23.49)	(3.16)	(.14)	(.67)	(.69)
5	27.90	12.04	32.00	13.97	.35	-.12	-.29
	(21.42)	(4.76)	(21.34)	(11.39)	(.21)	(1.32)	(.98)
7	17.82	12.30	19.72	12.76	.35	-.23	-.10
	(6.98)	(3.58)	(6.90)	(3.54)	(.26)	(.68)	(.70)
6	35.96	12.81	38.54	12.72	.17	.56	.35
	(21.55)	(4.58)	(27.29)	(3.60)	(.51)	(1.26)	(1.45)
8	24.38	12.25	27.50	11.32	.35	.06	.25
	(9.02)	(4.88)	(7.80)	(2.37)	(.22)	(.81)	(1.01)

**Note: Lower numbers indicate better performance*

Hypothesis 3: Males will score higher than females on the Mental Rotation Test with experience on spatial tasks as measured by demographic experience questions mediating the relationship between gender and scores on the MRT and on the experimental task

This hypothesis was tested in two ways: first, the mediation was tested using the method put forth by Barron and Kenny (1986). Several researchers (Field, 2013; Hayes 2009) have suggested that the Baron and Kenny method of evaluating mediation falls short in that it does not provide an estimate of the indirect effect of the predictor on the outcome when the mediator is present in the model. Because of this, a secondary analysis was performed using the *Process* function provided by Hayes (2009) in SPSS. This method allows for the estimation of the indirect effect of the outcome variable on the dependent variable with the proposed mediator present and does not rely on significance testing to find mediation.

The first step in the mediation analyses was to ensure that there was a relationship between the relevant variables. Table 14 shows the correlations between the experience variables being examined, gender, and mental rotation score (MRT score).

Table 14. Correlations of gender, spatial experience variables and MRT scores.

Measure	1	2	3	4	5	6
1. Gender						
2. Hours video games	-.378**					
3. First person perspective	-.674**	.511**				
4. Third person perspective	-.321**	.477**	.641**			
5. Visual arts	.185	.032	.066	.183		
6. Picture puzzles	.207*	-.056	.098	.264*	.223*	
7. Mental rotation score	-.305**	.288**	.342**	.217*	.073	.004

Note: $^{*}p<.05$, $^{**}p<.01$

Looking at the correlations between these variables revealed that: 1. Correlations between hours of video games, gender, and mental rotation score met the conditions necessary to test for possible mediation (on a bivariate level), 2. Correlations between 1st- person perspective games, gender, and mental rotation score met the conditions necessary to test for possible mediation, and 3. Correlations between 3rd- person perspective games, gender and MRT score met the conditions necessary to test for possible mediation. However, 4. Correlations between visual arts, gender, and MRT score and 5. Correlations between experience with picture puzzles, gender and MRT did not meet the conditions necessary to test for possible mediation and therefore these analyses were not performed. Only hours of video games played, experience with first person perspective games, and third person perspective games were considered for mediation analyses.

The first analysis focused on the relationship between hours of playing video games, gender, and mental rotation score. Table 15 shows the results of the three regressions that were used to test the mediation. Hours of playing video games partially mediated the relationship between gender and MRT performance. Gender was significantly related to MRT score, $F(1,90)=9.22$, $p=.003$, and the proposed mediator, hours of video games played per week (VG hours), $F(1,90)=14.98$, $p<.0001$. In the third analysis a hierarchical regression was performed. In the first step, VG hours was entered into the model as a predictor of MRT score. In the second block of the regression, gender was entered into the model as a predictor. In model 1, the relationship between VG hours and MRT score was significant $F(1,90)=8.16$, $p=.005$ and accounted for 8.3% of the variance. When gender was added, the overall model was significant, $F(2,89)=6.52$, $p=.002$ and accounted for 12.8% of the variance in MRT scores. The increase in R^2 ($\sqrt{R^2_{change}}=.21$) represented a small, but significant effect, $F(1,89)=8.16$, $p=.035$. In the second model, the relationship between VG hours and MRT remained significant, $\beta=.20$, $t=1.89$, $p=0.42$, while the relationship between gender and MRT score became weaker, $\beta=-.23$, $t=-2.14$, *p*=.04 compared to the direct relationship, *β*=.31 in model 1. These results support a partial mediation.

Table 15. Regression results and corresponding coefficients for the mediation analysis 1

	Regression Results			Coefficients	
	R	R^2	R^2_{Change}	B	β
Analysis One:					
MRT on gender	.31	.09**		-3.13**	-.31**
Analysis Two:					
VG hours on gender	.38	.14***		-4.7***	-.38***
Analysis Three:					
Model 1: MRT on VG hours	.29	.08**		.27**	29**
Model 2: MRT on VG hours	.36	.13**	.045*	.19*	.20*
MRT on Gender				-2.3*	-.23*

Note *=*p*<.05, **=*p*<.01, ***=*p*<.001

Using the Hayes (2009) *Process* tool in SPSS made it possible to test the indirect effect of gender on mental rotation score once VG hours were included in the model. This analysis showed that the indirect effect of gender on MRT score through video game hours per week was significant, B=-.78, BCa CI [-4.53, -.1655], R^2=.048. Figure 9 below shows a diagram of the mediation model.

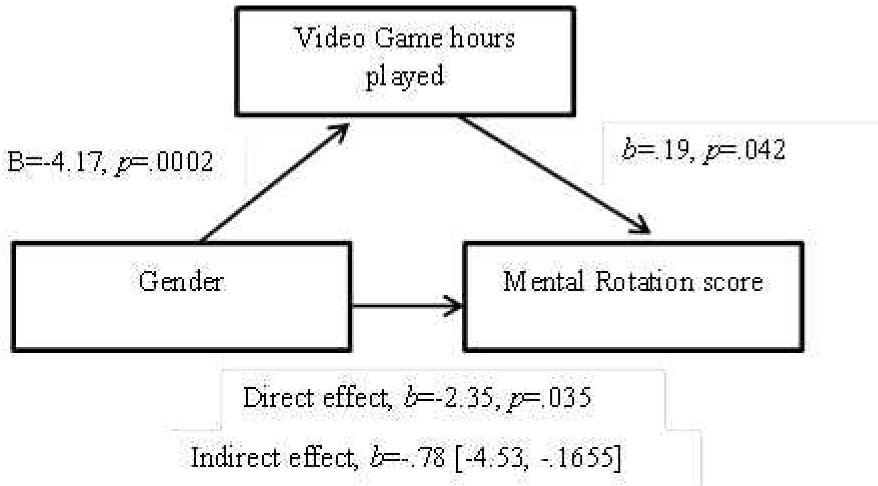


Figure 9: Diagram of the mediation model between gender, MRT score, and video game hours played per week

The second analysis focused on the relationship between gender, MRT score, and experience on first person perspective video games. The mediation was first analyzed according to the method proposed by Barron and Kenny (1986). Table 16 shows the results of the regression analyses and their corresponding coefficients. Hours of playing video games did not mediate the relationship between gender and MRT performance. Gender was significantly related to MRT score, $F(1,90)=9.22$, $p=.003$, and experience with first person perspective video games (FP games), $F(1,89)=74.03$, $p<.0001$. As with the previous analysis, a hierarchical regression was performed in order to verify mediation. FP games were entered in the first step as a predictor of MRT score. The relationship was significant $F(1,89)=11.76$, $p=.001$ and accounted for 12% of the variance in MRT scores. In the second step, both FP games and gender were entered into the model as predictors. The overall model was significant, $F(2,88)=6.32$, $p=.003$ and accounted for 13% of the variance in MRT scores although R^2_{change} was not significant, $F(1,89)=.90$, $p=.345$. When controlling for experience on first person games, the relationship of both predictors with MRT scores became non-significant, (FP games β=.26, t=1.39, p=0.62; Gender β=-.13, t=-.95, p=.345). This indicates that neither variable contributed a significant amount of unique variance when entered into the regression together.

Table 16. Regression results and coefficients for mediation analysis 2

	Regression Results			Coefficients	
	R	R^2	R^2_{Change}	B	β
Analysis One:					
MRT on gender	.31	.09**		-3.13**	-.31**
Analysis Two:					
FP games on gender	.67	.45***		-1.97***	-.68***
Analysis Three:					
Model 1: MRT on FP games	.34	.12***		1.2***	.34***
Model 2: MRT on FP games	.36	.13**	.009	.90	.26
MRT on gender				-1.3	-.13

Note *=*p*<.05, **=*p*<.01, ***=*p*<.001

The examination of the indirect effect of gender on MRT score through experience with 1st-person perspective games also indicated that mediation was not present, B=-1.77, Ba CI [-3.82, -.0059], R^2=.08. The direct effect of gender on MRT score was non-significant, meaning that the relationship between gender and MRT score became non-significant when controlling for experience on 1st- person shooters. Figure 10 shows the diagram of the mediation model.

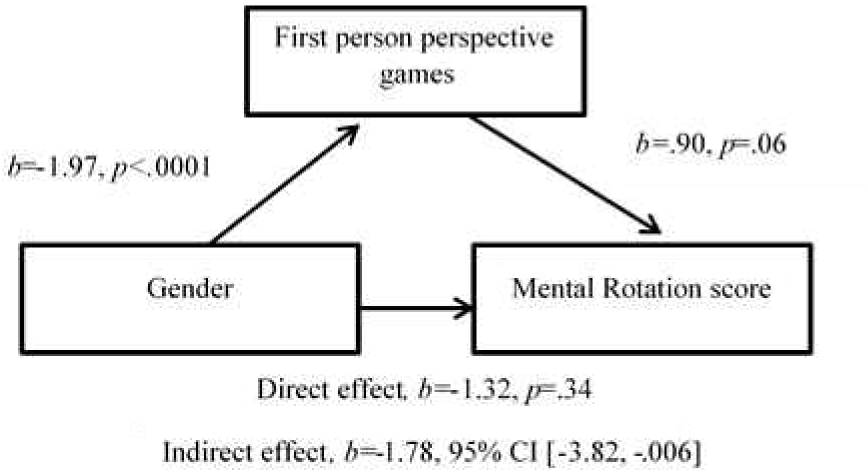


Figure 10: Diagram of the mediation model between gender, MRT score and 1st-person perspective video games

The last mediation analysis focused on the relationship between gender, MRT score, and experience on 3rd person perspective video games. Table 17 shows the results of the regression analyses and their associated coefficients. Experience on third person perspective games did not mediate the relationship between gender and MRT performance. Gender was significantly related to MRT score, $F(1,90)=9.22$, $p=.003$, and experience with third person perspective video games (TP games), $F(1,89)=10.24$, $p<.002$. Hierarchical regression was performed using FP game experience and gender as predictors. FP games were entered in the first step. The relationship was significant $F(1,89)=4.38$, $p=.039$ and accounted for 4.7% of the variance in MRT scores. In the second step, both FP games and gender were entered into the model as predictors. The overall model was significant, $F(2,88)=5.23$, $p=.007$ and accounted for 10.6% of the variance in MRT scores. The R^2_{change} was small but significant, $F(1,89)=.5.84$, $p=.018$. However, when gender was added as a predictor in model 2, the relationship between experience on TP games and MRT scores became non-significant, while the relationship between gender and MRT scores remained significant (TP games β=.13, t=1.26, p=.21; Gender β=-.26, t= -2.42, p=.018). This indicated that experience on third-person video games did not explain a significant amount of unique variance in MRT scores over and above gender.

Table 17. Regression results and coefficients for mediation analysis 3

	Regression Results			Coefficients	
	R	R^2	R^2_{Change}	B	β
Analysis One:					
MRT on gender	.31	.09**		-3.13**	-.31**
Analysis Two:					
TP games on gender	.32	.10**		-.87**	-.32**
Analysis Three:					
Model 1: MRT on TP games	.22	.05*		.83*	.22*
Model 2: MRT on TP games	.32	.09*	.06*	.51	.13
MRT on Gender				-2.66*	-.26*

Note *=p<.05, **=p<.01, ***=p<.001

Analyses were also performed using the *PROCESS* tool (Hayes, 2009) in SPSS in order to examine the indirect effect of gender on MRT score through experience on 3rd person video games. This analysis also indicated that no mediation was present, B=-.44 Ba CI [-1.51, .18], R^2=.03. When gender and 3rd person perspective video games were both entered into the predictive model for MRT, the relationship between 3rd person video games and MRT scores became non-significant. Additionally, the confidence interval for the indirect effect between gender and MRT through experience on 3rd-person perspective games included 0, indicating that it was not a true effect. A diagram of this relationship can be seen in Figure 11.

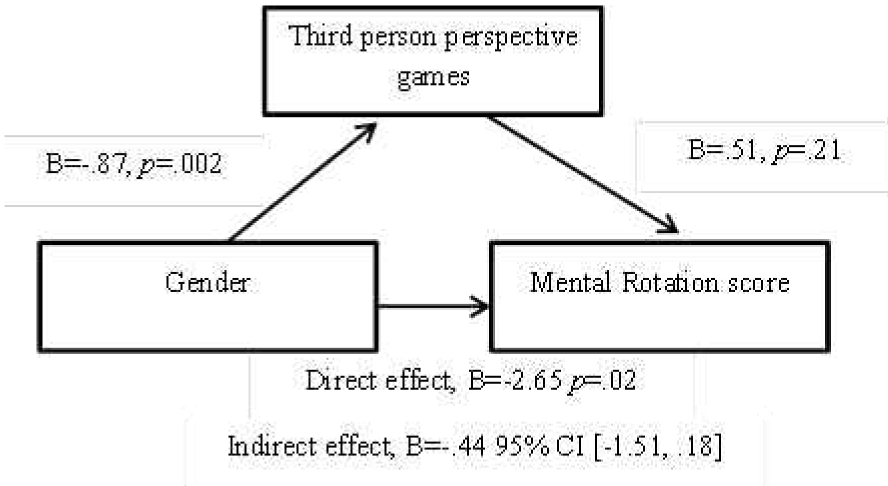


Figure 11: Diagram of the mediation model between gender, MRT score and 3rd-person perspective video games

A t-test revealed that males performed better in general on the MRT, M_{male}= 11.80 (5.40), M_{female}=8.72 (4.42), $t(92)=3.05$, $p=.003$. Further, a regression analysis showed that MRT scores were predictive of task performance, represented by post-test medians, $\beta=-.34$, $t(91)=-3.38$, $p<.001$ and explained a significant proportion of the variance in task performance $R^2=.113$, $F(1,90)=11.42$, $p=.001$. However, males did not perform better on the task. A t-test revealed that males did significantly better (M=33.29, SD=11.24) than females (M=41.09, SD=12.01), $t(90)=-3.16$, $p=.002$, $d=.67$ on pre-test and therefore pre-test was used as a covariate for further analyses. An ANCOVA was performed to compare the median post-test deltas of males and females and while the covariate was significant, $F(1, 89)=45.36$, $p<.0001$, partial $\eta^2=.338$, no significant differences were found on gender, M_{male}= 24.27 (15.31), M_{female}=26.24 (14.43), $F(1, 89)=1.96$, $p=.165$, partial $\eta^2=.022$. Similar analyses were performed for transfer median scores ($F(1, 88)=.68$, $p=.41$, partial $\eta^2=.008$) and pre to post-test gain scores ($F(1, 89)=1.63$, $p=.21$, partial $\eta^2=.018$). While males generally performed better than females on the post-test and transfer task, none of these differences were significant. Additionally, females attained higher gain scores from pre to post-test (See Table 18). During the first half of training, males outperformed females, $t(90)=-2.71$, $p=.008$, $d=.58$, however the difference between their medians in the second half of training was not significantly different, $t(90)=-1.60$, $p=.115$, $d=.34$.

The hypothesis that spatial experience would mediate the relationship between gender and MRT score was partially supported, however the theory that this would lead to better performance on the task was not.

Table 18. Means and standard deviations for performance by gender

	Post-test median*	Transfer median*	Pre to post-test gain	First half of training*	Second half of training*
Males	24.27 (15.32)	27.95 (19.32)	.28 (.26)	27.26 (12.03)	25.20 (15.66)
Females	41.09 (12.01)	30.81 (16.59)	.35 (.29)	35.70 (16.44)	30.34 (14.83)

**Note: Lower numbers indicate better performance*

Hypothesis 4: Participants in matched feedback conditions will perform better during training than participants in mismatched conditions (Groups 1, 2, 5, 6 > 3, 4, 7, 8)

A mixed-model repeated measures ANOVA was used to test the hypothesis that participants in matched conditions would perform better during training than participants in mismatched conditions. Every 15 scenarios during the training was considered a set, and as such there were 6 sets of scenarios. Scenario set was entered as the within subjects variable and Match/Mismatch was entered as the between subjects variable. Table 19 below shows the means and standard deviations for the matched and mismatched groups during training. The results of Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(14)=79.28$,

$p<.0001$. As such, the degrees of freedom for the F-tests for the between subject variable (set) were adjusted using the Huynh-Feldt estimates of sphericity, ($\varepsilon=.75$). There were significant main effect of set between the groups during training, $F(3.76,338.17)=7.836$, $p<.0001$, partial $\eta^2=.08$. The main effect of Match/Mismatch was also significant, $F(1,90)=7.79$, $p=.006$, partial $\eta^2=.08$ indicating the Mismatched group outperformed the Matched group during training. The interaction between set and Match/Mismatch was not significant, $F(3.75, 338.17)= .446$, $p=.76$, partial $\eta^2=.005$. According to within-subjects contrasts, there was a significant difference between Sets 1 and 2 ($F(1,90)=8.13$, $p=.005$, partial $\eta^2=.08$), Sets 3 and 4, ($F(1,90)=29.04$, $p<.0001$, partial $\eta^2=.24$), Sets 4 and 5, ($F(1,90)=18.11$, $p<.0001$, partial $\eta^2=.17$), and Sets 5 and 6, ($F(1,90)=9.07$, $p=.003$, partial $\eta^2=.09$). Figure 12 shows a graphical representation of the groups' median deltas as they progressed through training. The matched group's deltas tended to be higher (indicating poorer performance), the opposite of what was expected.

Table 19. Means and standard deviations for matched and mismatched groups during training.

Condition	Set 1*	Set 2*	Set 3*	Set 4*	Set 5*	Set 6*
Match	31.56	37.78	38.26	29.29	36.02	32.16
	(17.96)	(24.38)	(20.13)	(18.90)	(22.24)	(22.48)
Mismatched	25.21	31.40	28.98	22.85	26.89	22.32
	(10.15)	(20.52)	(13.16)	(9.67)	(12.16)	(17.57)

**Note: Lower numbers indicate better performance*

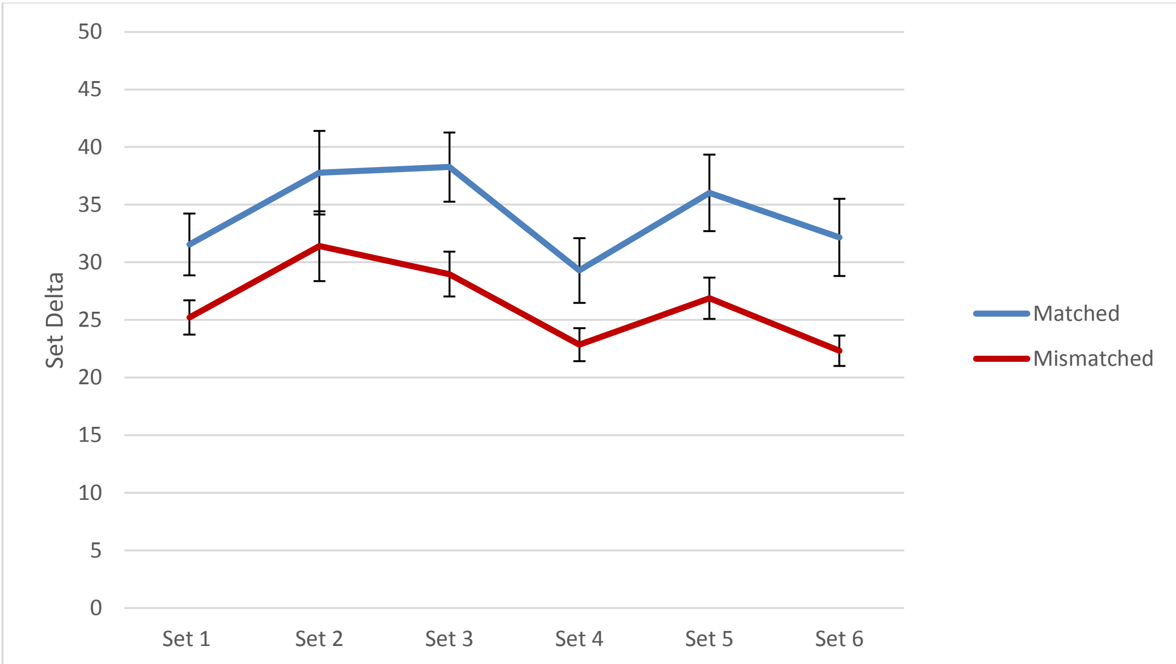


Figure 12: Matched and mismatched group deltas during training.

In order to examine differences between high and low spatial ability groups on the IV of match/mismatch, the Select Cases function in SPSS was used to isolate the high ability participants only. A mixed-model repeated measures ANOVA was used to examine the effects of the match/mismatch manipulation on higher ability participants. Table 20 shows the means and standard deviations for median deltas of these groups during training. The results of Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(14)=54.71$, $p<.0001$. As such, the degrees of freedom for the F-tests were adjusted using the Huynh-Feldt estimates of sphericity, ($\varepsilon=.65$). The analysis revealed a significant main effect of the within subjects variable of Set, $F(3.24, 142.48)=6.03$, $p<.0001$, partial $\eta^2=.10$. The contrasts showed that there was a significant difference between Set 1 and 2, ($F(1,43)=5.45$, $p=.024$, partial $\eta^2=.11$), Sets 3 and 4, ($F(1,43)=13.28$, $p=.001$, partial $\eta^2=.23$), and Sets 4 and 5, ($F(1,43)=5.35$,

p=.026, partial η^2=.11). The main effect of Match/Mismatch was not significant, $F(1,43)=1.12$, p=.37, partial η^2=.019, nor was the interaction between set and approach, $F(3.24, 142.48)=.51$, p=.69, partial η^2=.01. Figure 13 shows the visualization of the high ability matched and mismatched groups' median deltas during training.

Table 20. Means and standard deviations for the high ability matched and mismatched groups during training.

Condition	Set 1*	Set 2*	Set 3*	Set 4*	Set 5*	Set 6*
Match	23.64	29.23	29.84	22.50	26.59	23.05
	(10.13)	(19.18)	(14.46)	(17.51)	(17.84)	(16.07)
Mismatched	22.30	29.09	23.35	19.43	22.22	20.69
	(8.14)	(19.71)	(9.27)	(7.81)	(7.86)	(8.70)

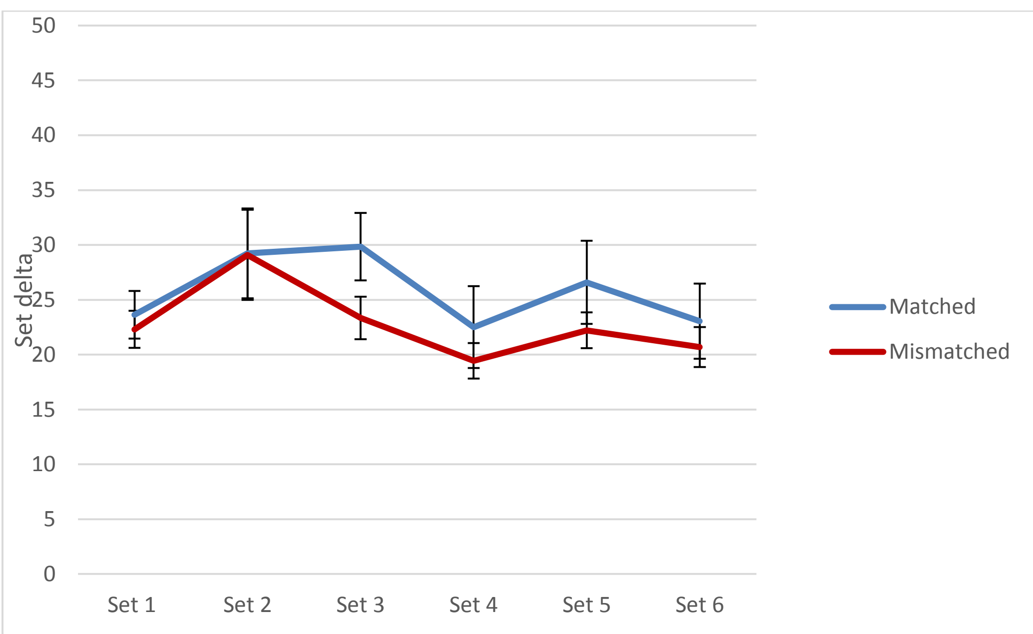


Figure 13: High ability matched and mismatched group deltas during training.

A similar analysis was performed for the lower ability participants. The means and standard deviations for the lower ability groups can be seen in Table 21. Once again, the results of the Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(14)=41.19$, $p<.0001$. The degrees of freedom for the F-tests were adjusted using the Huynh-Feldt estimates of sphericity, ($\varepsilon=.81$). The analysis revealed a significant main effect of the within subjects variable of Set, $F(4.02, 177.20)=3.40$, $p=.01$, partial $\eta^2=.07$. Contrasts revealed significant differences between Sets 3 and 4, ($F(1,45)=16.51$, $p<.0001$, partial $\eta^2=.27$), Sets 4 and 5, ($F(1,45)=12.91$, $p=.001$, partial $\eta^2=.22$), and Sets 5 and 6, ($F(1,45)=7.02$, $p=.011$, partial $\eta^2=.14$). The main effect of Match/Mismatch was also significant, $F(1,44)=10.71$, $p=.002$, partial $\eta^2=.12$. The mismatched groups outperformed the matched groups during training. However, the interaction between set and Match/Mismatch was not, $F(4.02, 177.20)=.51$, $p=.73$ partial $\eta^2=.01$. Figure 14 shows the visualization of the low ability matched and mismatched groups' median deltas during training. The graph shows that the mismatched feedback approach was more beneficial for lower ability participants, the opposite of what was hypothesized.

Analyses were also performed for all participants to assess the difference between the first half and second half of training. The mean of the median deltas for the first three sets was used to create a variable that represented the first half of training, and the mean of the median deltas for the last three sets of the training session was used to create a variable that represented the second half of training. A t-test was performed to determine if there were any differences between the matched and mismatched groups on this variable. There was a significant difference between the groups for the first half of training, $t(72.14)= -2.34$, p=022, equal variances not assumed, where the mismatched group (*M*=28.53, *SD*=11.01) outperformed the matched group. A similar result was found for the second half of training, $t(60.23)=-2.70$, $p=.009$, equal variances not assumed, where the mismatched group (*M*=24.02, *SD*=8.58) outperformed the matched group (*M*=32.49, *SD*=19.26).

Hypothesis 4 was not supported.

Table 21. Means and standard deviations for the lower ability matched and mismatched groups during training

Condition	Set 1*	Set 2*	Set 3*	Set 4*	Set 5*	Set 6*
Match	39.13	45.95	46.32	35.78	45.04	40.86
	(20.60)	(26.35)	(21.72)	(18.22)	(22.61)	(24.53)
Mismatched	28.00	33.63	34.38	26.13	31.38	23.88
	(11.22)	(21.45)	(14.22)	(10.28)	(13.93)	(9.00)

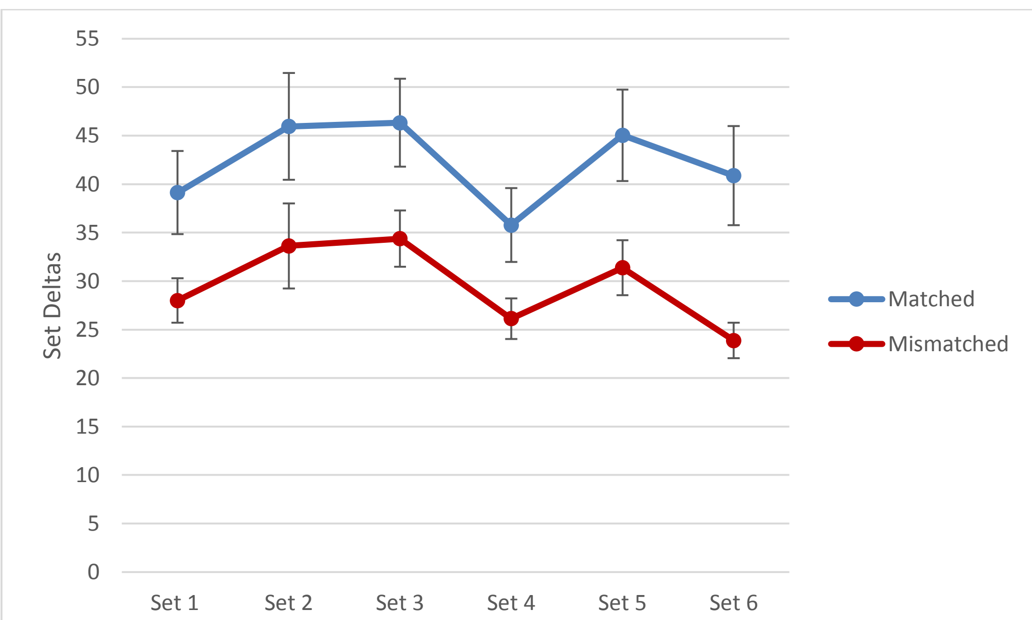


Figure 14: Lower ability matched and mismatched group deltas during training.

Hypothesis 5: Participants in the hybrid conditions will perform better during training than participants in the ATI conditions (Groups 1-4 > 5-8)

In order to examine how AT Approach affected training for people of high and low spatial ability during training separately, the Select Cases function in SPSS was used to isolate the high ability participants only. A mixed-model repeated measures ANOVA was used to examine if the Hybrid method of adaptive training was beneficial for high ability participants during training. The means and standard deviations for high ability participants can be seen in Table 22. Once again, the results of Mauchly's test indicated that the assumption of sphericity had been violated for the within subjects variable, $\chi^2(14)=60.37$, $p<.0001$. The degrees of freedom for the F-tests were adjusted using the Huynh-Feldt estimates of sphericity, ($\varepsilon=.62$). The analysis revealed a significant main effect of Set, $F(6.17, 137.20)=6.17$, $p<.0001$, partial $\eta^2=.12$. Contrasts revealed significant differences between Sets 1 and 2, ($F(1,43)=5.51$ $p=.024$, partial $\eta^2=.11$), Sets 3 and 4, ($F(1,43)=13.46$, $p=.001$, partial $\eta^2=.24$), and Sets 4 and 5, ($F(1,43)=6.48$, $p=.015$, partial $\eta^2=.13$). There was no significant main effect of Approach, $F(1,44)=.62$, $p=.44$, partial $\eta^2=.01$. The interaction between set and approach was also not significant, $F(3.118, 137.20)=2.01$, $p=.098$, partial $\eta^2=.04$. Figure 15 shows the visualization of the high ability participants' performance during training.

Table 22. Means and standard deviations of the high ability group deltas during training

Condition	Set 1*	Set 2*	Set 3*	Set 4*	Set 5*	Set 6*
Hybrid	23.00	28.43	25.93	17.61	25.65	19.04
	(9.04)	(15.75)	(10.92)	(8.06)	(11.45)	(8.89)
ATI	23.30	32.09	28.04	24.30	22.65	24.57
	(9.32)	(24.49)	(14.35)	(16.44)	(15.60)	(15.18)

**Note: Lower numbers indicate better performance*

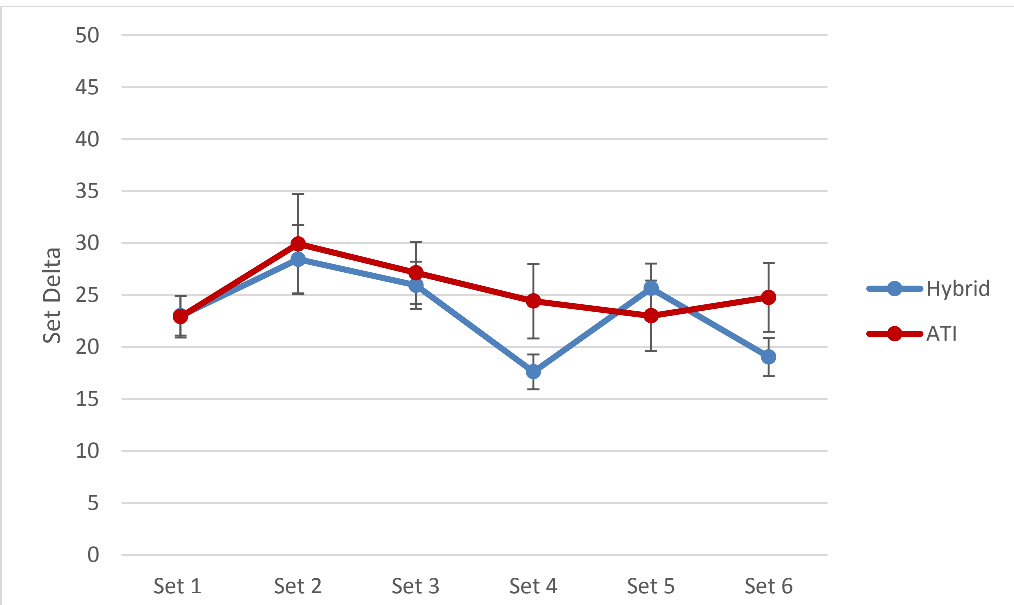


Figure 15: High ability ATI and Hybrid group median deltas during training

These analyses were not performed for all eight groups, or the low ability Hybrid and ATI groups, because as mentioned previously, the low ability Hybrid group received the same training manipulation as the low ability ATI group. Therefore, these analyses could not be included.

Hypothesis 5 was not supported.

Exploratory analyses

Several exploratory analyses were performed. Firstly, participant scores on the self-efficacy questions were examined to see if participants in matched and hybrid conditions rated their self-efficacy on the task higher than participants in mismatched or ATI conditions.

Self-Efficacy

There were no significant differences between the groups on the self-efficacy individual questions or on total self-efficacy score: Accurately find AOB, $F(7,84)=1.05$, $p=.40$, $\eta^2=.08$; Quickly find AOB, $F(7,84)=1.66$, $p=.13$, $\eta^2=.12$; Find AOB at night, $F(7.84)=1.12$, $p=.36$, $\eta^2=.08$; Find AOB in high sea state, $F(7,84)=.68$, $p=.69$, $\eta^2=.05$; Total self-efficacy, $F(7,84)=.98$, $p=.45$, $\eta^2=.08$. The means for the self-efficacy questions can be seen in Table 23.

Table 23. Means and standard deviations of self-efficacy questions

Condition	Accurately find AOB	Quickly find AOB	Find AOB at night	Find AOB in high sea state	Total Self efficacy
1	3.27 (.65)	3.00 (1.09)	2.09 (.94)	3.09 (1.14)	2.86 (.85)
2	3.30 (.95)	3.30 (1.16)	2.30 (.95)	3.30 (1.16)	3.05 (.88)
3	3.33 (.78)	3.10 (1.03)	1.92 (.79)	2.75 (.96)	2.79 (.68)
4	3.04 (.75)	3.00 (1.04)	1.67 (.49)	3.25 (1.14)	2.74 (.75)
5	2.82 (.75)	3.36 (1.21)	1.90 (.94)	3.36 (1.36)	2.86 (.93)
6	2.69 (1.03)	2.62 (.87)	1.54 (.66)	2.69 (.95)	2.38 (.72)
7	3.27 (.47)	3.91(.70)	1.82 (.60)	3.27 (.65)	3.07 (.42)
8	3.08 (.90)	2.83 (1.11)	1.75 (.62)	2.92 (1.24)	2.64 (.85)

Note: 1=Not confident at all, 5=very confident

High Versus Low Performers

In the next analysis, a median split was performed on post-test median deltas to examine if high performers benefitted more from the feedback manipulations or the adaptive approach manipulations. The median for post-test score is 21.5, SD=14.76. The means and standard deviations for the high performers can be seen in Table 24. There were no significant differences between high performers who received hybrid AT and high performers who received ATI AT on post-test median, $t(38.62)=1.50$, $p=.141$, $d=.45$ (equal variances not assumed), on transfer median, $t(43)=.29$, $p=.775$, $d=.08$, or on gain, $t(31.76)=-1.37$, $p=.179$, $d=.40$ (equal variances not assumed).

There were also no significant differences between high performers who received matched feedback and high performers who received mismatched feedback on post-test median, *t*(41.46)=-1.31, *p*=.20, *d*=-.40 (equal variances not assumed), transfer median, *t*(43)=-1.12, *p*=.27, *d*= -.35, and gain, *t*(43)=.21, *p*=.84, *d*=-.07.

Table 24. Means and standard deviations of high performers on the IVs Approach and Feedback type

	Post-test median*	Transfer median*	Pre to post-test gain
ATI	38.50 (18.06)	38.35 (23.89)	.11 (.37)
Hybrid	31.64 (12.12)	36.46 (19.93)	.23 (.17)
Matched	37.74 (18.03)	40.68 (22.89)	.16 (.36)
Mismatched	31.90 (11.73)	33.35 (20.22)	.18 (.19)

**Note: Lower numbers indicate better performance*

Similar analyses were performed for participants who were considered low performers on the post-test. Their means and standard deviations are shown in Table 25. There were no significant differences between low performers (on post-test) who received matched vs mismatched feedback on post-test median, *t*(45)=.23, *p*=.82, *d*=.06, transfer median, *t*(44)= -1.57, *p*=12, *d*=-.47, or gain, *t*(45)=-.79. *p*=.43, *d*=.23. Analyses were not performed on the lower ability participants for the variable of AT Approach because the lower ability Hybrid group did not receive different training manipulations than the lower ability ATI group.

Table 25. Means and standard deviations of low performers on the IV of Feedback type

	Post-test median*	Transfer median*	Pre to post-test gain
Matched	15.92 (3.79)	23.65 (6.66)	.50 (.15)
Mismatched	16.19 (3.97)	20.58 (6.50)	.47 (.16)

NASA-TLX

For the next analysis, the remaining scales of the NASA TLX were examined. In addition to Mental Demand, participants filled out ratings for demands on Time, Performance, Effort, and Frustration. Again, these ratings were given once after the training portion of the study and once after the post-test. There were no significant differences between the groups on any of these ratings at Time 1, Time: $F(7,84)=2.03$, $p=.06$, $\eta^2=.14$; Performance: $F(7,84)=1.40$, $p=.22$, $\eta^2=.10$; Effort: $F(7,84)=1.04$, $p=.41$, $\eta^2=.08$; Frustration: $F(7,84)=1.85$, $p=.09$, $\eta^2=.13$. No significant differences were found on these variables at Time 2, Time: $F(7,84)=1.30$, $p=.26$, $\eta^2=.10$; Performance: $F(7,84)=.38$, $p=.91$, $\eta^2=.03$; Effort: $F(7,84)=1.89$, $p=.08$, $\eta^2=.14$; Frustration: $F(7,84)=1.16$, $p=.34$, $\eta^2=.09$. The means and standard deviations for these measures can be seen in Tables 26 and 27.

Table 26. Means and standard deviations for NASA TLX ratings at Time 1

Condition	Temporal Demand	Performance	Effort	Frustration
1	33.64 (24.34)	36.09 (18.94)	62.00 (19.32)	46.09 (21.82)
2	45.70 (27.56)	51.90 (26.82)	73.10 (22.19)	39.10 (30.26)
3	28.83 (19.84)	38.08 (14.94)	67.00 (16.87)	29.50 (17.87)
4	48.83 (30.18)	48.83 (25.42)	62.00 (24.41)	44.50 (25.21)
5	46.18 (26.78)	41.82 (21.04)	74.45 (17.39)	63.91 (25.67)
6	32.31 (22.78)	58.00 (25.31)	65.31 (12.19)	46.77 (30.68)
7	19.18 (14.14)	46.55 (17.24)	58.91 (16.74)	35.64 (17.26)
8	45.83 (30.92)	44.17 (16.04)	60.17 (19.78)	51.58 (32.85)

Table 27. Means and standard deviations for NASA TLX ratings at Time 2

Condition	Temporal Demand	Performance	Effort	Frustration
1	25.73 (22.87)	43.64 (17.73)	64.55 (22.67)	39.45 (22.89)
2	32.90 (25.92)	40.20 (20.93)	68.10 (24.42)	28.30 (27.66)
3	27.50 (18.94)	45.50 (19.22)	61.42 (23.54)	31.42 (22.07)
4	41.42 (22.90)	44.58 (27.00)	46.83 (26.80)	31.42 (19.63)
5	24.55 (22.18)	43.82 (24.17)	61.54 (27.07)	43.09 (20.64)
6	26.15 (14.29)	52.46 (23.11)	51.31 (24.36)	28.85 (27.87)
7	20.64 (19.37)	41.00 (18.64)	40.55 (16.66)	26.36 (21.87)
8	38.67 (26.22)	44.42 (14.15)	51.58 (17.51)	46.08 (23.39)

Working Memory

Next, participants' performance on the working memory measure was examined. An ANOVA was performed to determine if there were differences between the groups on this measure; no significant differences were found, $F(7, 84)=1.51$, $p=.17$, $\eta^2=.13$. The means for each group on the working memory measures can be found in Table 28.

Table 28. Mean scores on the working memory measures

Condition	Working memory score
1	3.00 (.97)
2	2.65 (1.37)
3	3.04 (.96)
4	2.62 (1.13)
5	2.59 (1.43)
6	1.88 (1.06)
7	3.23 (1.27)
8	2.58 (1.12)

In order to examine the effect of spatial ability on working memory, the groups were collapsed on the variable of spatial ability to create a High and Low spatial ability group. This variable was entered into a t-test as the independent variable with working memory scores as the dependent variable. A significant relationship was found, $t(90)=-2.27$, $p=.026$, $d=.47$ indicating that higher ability participants achieved higher working memory scores ($M=2.96$, $SD=1.15$) than lower ability participants ($M=2.41$, $SD=1.17$) .

Next, a regression analysis was calculated using working memory score as the predictor variable and post-test median scores as the as the dependent variable. The relationship, $\beta=-.35$, $t=-3.63$, $p<.0001$, was significant, $F(1, 90)= 13.18$, $p<.0001$ with working memory scores predicting 12.8% of the variance. Earlier, I found that spatial ability (represented by MRT scores) predicted 11.3% of the variance in post-test median scores. In order to determine if working memory could predict variance above that of spatial ability, a hierarchical regression was performed. In the first step, spatial ability (MRT scores) was entered as a predictor of post-test median scores. In step two, working memory scores were added as a predictor. As found previously, in model 1 the relationship between spatial ability and post-test median scores was significant (recall β=-.34, $t(91)$ =-3.38, p<.001, R^2=.113, $F(1,90)$=11.42, p=.001). In model 2, regression equation was significant, $F(2, 89)$= 10.19, p<.0001. This model predicted 18.6% of the variance in post-test median scores. The change statistics were also significant, R^2_{change}=.07, $F(1,89)$=8.07, p=.006. The regression coefficients and their associated significance tests can be found in Table 29. These results suggest that both working memory capacity and spatial ability contributed significant unique variance in predicting post-test performance.

Table 29. Regression coefficients for spatial ability and working memory scores as predictors of post-test performance

	B	*β*	*t*	*p*
Model 1:				
MRT score	-.97	-.33	-3.38	.001
Model 2:				
MRT score	-.73	-.25	-2.53	.013
Working memory score	-3.51	-.28	-2.84	.006

Call Times During Training

Trends in the data suggested that groups who performed better during training or on the post-test had higher call times. Perhaps this trend in data was an indication of the effort participants' put into making their calls accurately. Analyses were performed to determine if participants who took longer to make their calls during training performed better on the task.

A median split was used to divide the participants into tertiles based on the mean time it took them to complete calls during training. The median scenario time, calculated by averaging all of the participants calls over training, was 11.95 (4.22) seconds and the median post-test time was 11.85 (3.82) seconds. The split created a High, Middle, and Low group based on call times where the Low group took the shortest amount of time to make calls and the High group took the longest. Post-test performance was examined first; the means and medians for each group can be seen in Table 30. An ANCOVA was performed using pre-test median scores as the covariate. Pre-test median was significantly related to post-test median scores, $F(1, 88)= 50.07$. $p<.0001$, partial $\eta^2=.36$. Additionally, the time it took participants to complete their calls during training was significantly related to post-test scores, $F(1,88)= 5.38$, $p=.006$, partial $\eta^2=.11$. Post hocs (LSD) revealed that the Low group's post-test median scores were significantly worse than that of the Middle group ($M_{diff}=8.25$, $p=.029$). While the difference between the Low and High group was not significant when the significance was two-tailed ($M_{diff}=6.48$ $p=.084$), it was significant when one-tailed significance testing was used ($p=.042$) showing that the Low group performed worse than the High group. There were no significant differences between the Middle and High groups.

Similar analyses were completed for participants' transfer median scores. While the covariate (pre-test) was found to be significantly related to transfer median, $F(1,87)= 25.09$, $p<.0001$, partial $\eta^2=.22$, the time it took participants to complete calls during training was not, $F(1,87)= .57$, $p=.57$, partial $\eta^2=.013$.

Next, an ANOVA was used to examine pre to post-test gain scores. There was a significant difference between the groups, $F(1,88)= 5.61$, $p=.005$, partial $\eta^2=.11$. Post hoc (LSD) tests revealed that the Low group performed significantly worse than the Middle and High groups, $M_{diff}=-.19$ $p=.006$, $M_{diff}=-.21$ $p=.003$ respectively. There was no significant difference between the High and Middle groups.

These results indicate that participants who took more time to complete their calls during training performed better on the post-test and achieved higher gain scores than participants who made their calls faster.

Table 30. Means and standard deviations for task performance based on median call time during training

	Post-test median*	Transfer median*	Pre to post-test gains
Low	30.37 (16.86)	31.00 (16.13)	.19 (.34)
Middle	22.11(13.61)	26.19 (13.69)	.38 (.22)
High	23.89 (12.77)	29.28 (17.77)	.33 (.28)

**Note: Lower numbers indicate better performance*

Finally, ANCOVAs were performed to determine if higher or lower ability participants took longer to make their calls than lower ability participants. The means and standard deviations call times during training, post-test, and transfer can be seen in Table 31. Pre-test times were used as a covariate for the analyses below because it was determined that there was significant difference between the pre-test call times of the lower and higher ability participants, $t(90)=-1.76$, $p=.041$, one-tailed.

With pre-test times as a significant covariate $F(1,89)= 116.07$, $p<.0001$, partial $\eta^2=.57$, there was no significant relationship between ability and training call time, $F(1,89)= .59$, $p=.44$, partial $\eta^2=.007$. A similar relationship was found for post-test call times: Pre-test time $F(1,89)= 44.98$, $p<.0001$, partial $\eta^2=.34$; spatial ability $F(1,89)= .009$, $p=.92$, partial $\eta^2<.0001$. There was no significant relationship between spatial ability and transfer task call time, $F(1,88)= .39$, $p=.53$, partial $\eta^2=.004$. Although the covariate was significant, $F(1,88)= 67.08.57$, $p<.0001$, partial $\eta^2=.43$.

Table 31. Means and standard deviations for the call times of the higher and lower ability groups

	Pre-test	Training	Post-test	Transfer
High	20.57 (7.38)	13.68 (4.17)	12.75 (3.65)	13.57 (6.41)
Low	18.25 (5.09)	12.08 (4.17)	12.00 (3.98)	11.76 (3.29)

CHAPTER FOUR: DISCUSSION

Generally, the results of this study did not support the use of CLT-derived feedback adaptations based on spatial ability (for both higher and lower ability participants), or the use of Hybrid adaptive training (for higher ability participants). Below I will discuss the specific hypotheses and explanation of the results.

The Effect of Match/Mismatch

The first hypothesis that the groups who received feedback matched to their spatial ability would outperform those who receive mismatched feedback was not supported. Counter to expectations, the higher spatial ability participants who received more feedback during training outperformed those who received less feedback on the post-test, transfer, and on learning gains, although these results were not significant (See Appendix L). Similarly, the performance of the lower spatial ability participants was contrary to the hypothesized relationship; lower ability participants who received less feedback outperformed those who received more feedback on the post-test, transfer task, and on learning gains. Similar results were found for Hypothesis four which surmised that participants who received feedback matched on their spatial ability would perform better during training than participants who received mismatched feedback. This hypothesis was also not supported. When looking at all of the groups together, there was a significant main effect of the Match/Mismatch variable where participants who received feedback mismatched to their skill performed better during training, the opposite of what was expected. When looking at the higher ability trainees, there was no significant difference between those who were matched or mismatched during training.

However the trends suggest that those who were mismatched and received more feedback performed slightly better during training than those who received less feedback (See Table 10, Appendix L). For the lower spatial ability participants, there was a significant main effect of Match/Mismatch, which showed that the participants who received less feedback during training (mismatched) performed better.

Previous research (Landsberg, et al., 2012) had shown that using CLT-based adaptation could be effective when adapting based on performance. When participants were performing well they received less feedback, and when they were performing poorly they received more. The goal of the current research was to examine if this idea could be extended to include ability, but this theory was not supported. Although I was able to show that spatial ability was related to task performance (participants with higher spatial ability performed better than those with lower ability), the amount of feedback participants received made no difference on how they performed. Mental effort ratings as well as instructional efficiency scores indicated that the manipulation of matching or mismatching feedback based on spatial ability using the theory of CLT expertise reversal effect did not have the intended effect; matched groups did not report decreased subjective mental effort and their instructional efficiency scores for both outcomes and learning were not higher than those of participants in mismatched feedback groups. The literature on CLT suggests that expertise on a task is developed when trainees combine individual pieces of information that would initially be manipulated in working memory in order to create schemas that can be stored in long-term memory. Consequentially, these schemas reduce the amount of cognitive load associated with task performance because they free up working memory space (van Merriënboer & Sweller, 2005). The analyses on working memory showed that the higher ability participants generally showed a larger working memory capacity and therefore had more resources to use during the task. Perhaps because the task was complex, and because it was novel for all participants (i.e. there were no experts), the higher ability participants performed better when they were given more feedback because they had the requisite resources to process it. The more detailed feedback then would have helped them to improve on the task, while the less detailed feedback may not have contained enough information to improve performance. On the other hand, it is possible that lower spatial ability participants were already utilizing most of their cognitive resources to perform the task, leaving no processing resources for attending to and integrating the more detailed feedback. Perhaps the less detailed feedback helped them improve their performance without overwhelming their cognitive resources and increasing extraneous load to a point where performance was negatively affected. These results are contrary to the findings of Kelley and McLaughlin (2012) mentioned previously and do not support their finding that participants that are high in a task-related ability require less feedback. It is also possible that the process feedback tips were not helpful for the lower ability participants because they did not pinpoint the participants' specific problems. This may have created more extraneous load, rather than reducing it. Although the feedback statements were chosen to cover the most likely errors for each scenario, it is possible that they did not address the particular error the participant made. If this were the case, it may have been more detrimental to the lower spatial ability participants.

The working memory results may also help explain why higher spatial ability participants performed better on the task in general, regardless of the feedback they received. Consistent with Shah and Miyake (1996), higher spatial ability participants also exhibited higher levels of working memory capacity. Therefore, they may have been better at holding mental images of the ships in memory while performing the mental rotation necessary to determine the angle of each contact. However, contrary to the findings in Shah and Miyake (1996), no mediation was found between spatial ability, working memory, and performance. When entered into a hierarchical regression, both spatial ability and working memory contributed significant unique variance to the model (see Table 29).

The Effect of AT Approach

The second hypothesis stated that participants who received the Hybrid AT would outperform those who received the ATI AT. This hypothesis was not supported. Similarly, the fifth hypothesis, in which participants who received Hybrid AT would perform better during training than those that received ATI AT training, was not supported. Looking at the trends (See Appendix L) for the higher ability participants revealed that the Hybrid approach did lead to better performance on the post-test and gain scores, but not on transfer task performance. Additionally, those who received the Hybrid AT performed better during training, but not significantly so. When looking at the higher ability participants' performance during training, there was no significant difference between those who received Hybrid and ATI training, although those who received Hybrid training performed better in general, with the exclusion of set 5 where participants who received ATI performed better. It is possible that the failure to find significant differences for between the different AT approaches was due to the large amount of variance in the data which may have decreased the power and made it difficult to detect differences between the groups (See Tables 12 and 20). Another possibility is that the feedback manipulation in the Hybrid was not sensitive enough to create larger differences between the groups. Specifically, during the last half of training (or the last 45 scenarios), the feedback adapted based on performance every set, i.e. every 15 scenarios. This means that the participants in the hybrid condition had only three chances to receive different feedback based on their performance.

As stated previously, the lower ability participants could not be compared on this variable, because there was very little difference in the feedback received by the Hybrid and ATI groups (See Table 6). For the lower ability participants in the hybrid conditions, the cut-offs chosen for performance may have been too high; For the lower ability Hybrid Matched group, the participants' performance was never adequate enough (i.e. their median deltas were never low enough) to receive anything but the more detailed feedback (with the exception of two participants), making that group nearly equivalent to the lower ability ATI Matched group. Similarly, in the lower ability Hybrid Mismatched group, their performance was never high enough (with the exception of two participants) to receive the more detailed feedback, making them equivalent to the lower ability ATI Mismatched group. The cut-off scores were determined from previous studies that utilized the same testbed. However, there were differences between the above-mentioned studies and the current one that may have changed the effectiveness of the cut-off score chosen. In the previous study (Landsberg, et al., 2012), participants were given adaptive difficulty in addition to adaptive feedback. Further, the feedback adaptations occurred based participant performance on a single scenario, rather than a set of 15. Lower ability participants in the study began training at an easier task level, and moved through difficulty levels based on their performance on the task. The adaptive difficulty may have been more beneficial to lower ability participants who had less working memory capacity to perform the task. More specifically, it is possible that starting lower ability participants with easier scenarios decreased the intrinsic load of the task which allowed them to integrate feedback more effectively early in training. Moreover, providing feedback on a case-by-case basis rather than basing the adaptation on the performance of multiple scenarios may have decreased extraneous load. The combination of these two factors may have contributed to the difference in the effectiveness of the cut-off scores between the two studies.

The results of the current study indicate that the training provided never allowed the lower ability participants in these groups to attain a median delta lower than 30° in the last three sets of training. Several factors could have contributed to this issue. First, as mentioned above, lower ability participants may have benefited from an adaptive difficulty manipulation to decrease the intrinsic load of the task early in training. Lower ability participants did worse on the task in general. As mentioned previously this may have been due to differences in working memory that made mental rotation more difficult and require more resources. It is possible that lack of participant motivation may have contributed to performance during the task.

Experience on Spatial Tasks

Lastly, the hypothesis that experience on spatial tasks would mediate the relationship between gender and MRT performance was partially supported. Specifically, the total number of hours participants reported playing video games partially mediated the relationship between gender and score on the MRT. However, no mediation was found between gender, first person video game experience, third person video game performance, and MRT scores although gender differences were found between these variables. This may have been the result of a range restriction for the variables of first and third person video game experience as these variables were rated on a Likert scale rather than on an interval scale as was hours of video games played. This may have reduced the covariance between the variables, reducing the possibility of finding an effect. For first person video game experience, it was found that when both variables were included in a hierarchical regression, neither contributed unique variance in predicting scores on the mental rotation test. Also, when controlling for gender, it was found that the relationship between third- person video games and MRT scores were not significant. Further, the suggestion that differences in gender would lead to better task performance was not supported. Males did in fact attain higher scores on the MRT, but they did not perform significantly better on the task than females, even though MRT scores were shown to predict task performance. While males performed generally better on the post-test and transfer tasks, females attained higher gain scores from pre to post-test (because females were poorer performers on pre-test, they had more to gain from the training than males; see Table 17). As stated above, spatial ability was predictive of task performance; however, this variable only predicted 11% of the variance in MRT scores. When paired with the scores on the working memory measure, the combination of these variables predicted 18% of the variance, each contributing unique explanation for variance in performance. These results can be illustrated by the relationships found between gender, spatial ability, and working memory. A significant difference was found between higher spatial ability participants and lower spatial ability participants on working memory; however, no differences were found between males and females on working memory scores. These results suggest another source of variance (besides working memory) may have been responsible for the spatial ability differences between the males and females in the group.

Limitations and Future Research

Several limitations may have led to the findings in this study. First, the manipulations of the independent variables may not have been strong enough to create differences between the groups. Specifically, for the feedback manipulations, the higher feedback condition was a mix of both process and outcome feedback while the lower feedback condition consisted of outcome feedback. The effects may have been stronger if the conditions were process alone and outcome alone. The results found in this study are contrary to a previous finding (Kelley & McLaughlin, 2012) that participants who were higher in a task-related ability performed better when they were given outcome feedback and did better on a task not related to their ability when given a combination of process and outcome feedback. However, the tasks and abilities chosen were different in the study mentioned above. Future research should consider the interaction between the complexity of the task being trained, the abilities chosen, and how adaptation should occur based on those abilities.

As mentioned above, the cut-off criterion for lower ability participants was too high, which caused the two Hybrid conditions to be near identical to the two ATI conditions for lower ability participants. Although these cut-offs were useful for the higher ability participants, and while the cut-off score was chosen based on previous research using the same testbed, there were several differences between the studies that may have been vital in determining the utility of the cut-off chosen, particularly for the lower ability participants. Follow-on research to this study should consider using a lower cut-off for performance, or alternatively, adding an adaptive difficulty component to increase the performance of lower ability participants. This would make it possible to examine the utility of a Hybrid approach versus an ATI approach for lower ability trainees.

One interesting extension of this study would be examining if certain adaptations would be more useful to lower ability participants than higher ability participants. I was not able to determine from the current research if either of the adaptive approaches was more useful to one group than the other; the higher ability group performed better on the training, post-test, transfer, and achieved higher gain scores regardless of the experimental manipulations. However, future research could examine if lower spatial ability participants could reach a comparable level of performance of higher ability trainees when given different combinations of feedback adaptations, adaptive difficulty, and approaches.

As an extension of the current research, future research could examine the scheduling of adaptive training. As mentioned above, the Hybrid group participants in this study received either more or less feedback based on their performance on a set of 15 scenarios. In the current study this was not found to be significantly more effective than receiving feedback based on spatial ability. However, previous research found positive results when feedback was adapted based on participants' performance on each scenario (Landsberg et al., 2012). Future studies may examine this relationship by comparing groups who receives adaptations based on a single point of performance to a group who receives adaptations based on multiple points of performance, In this study, I was not able to find a difference between Hybrid and ATI adaptive training for the high ability participants. While the trends were in the correct direction, the findings were not significant. The results of previous research (Landsberg et al., 2012) suggested that increasing the sensitivity of the performance-based adaptation in the last half of training may increase the chance of finding significant results as this may make the difference between the Hybrid and ATI training more pronounced.

The current research also showed a strong connection between performance on the pre-test and achievement during training, the post-test, transfer task, and on gain scores. This suggests there may be a benefit to basing adaptations on pre-test performance rather than spatial ability. This could be useful in both an ATI and a Hybrid approach. In an ATI approach, the pre-test performance would determine the feedback throughout training, while in a Hybrid approach it would determine the adaptation in the beginning of training, followed by current performance-based training later. Possibly, another ability, or combination of abilities would have been a more useful basis for adaption. While there was a relationship between spatial ability and performance on the task, the results of the study suggest there may have been other factors that contributed to performance on the task. Future research might explore adapting training based on a combination of abilities or trainee attributes such as spatial ability, motivation, and working memory capacity. For instance, research has suggested that motivation increases the capacity of working memory (Paas, Renkl, & Sweller, 2003; van Merriënboer & Sweller, 2005). It follows then that those participants who are more motivated to learn a task will invest more effort in learning and perform the task better. Further, this connection may particularly important for tasks that are complex when intrinsic load is high. One research question to explore is whether lower ability participants who are more highly motivated to complete the task would perform similarly to higher ability participants who are not motivated.

Related to this idea, analyses indicated the possibility of Christmas treeing, or a lack of effort that may have led to lower scores during training and on the post and transfer tests as participants who took longer to make their calls during training performed better on the task. This result also indicates there may have been low motivation for participants to perform the task well. In future research, it may be useful to utilize a military population, who may have more intrinsic motivation to perform well on the task. Alternatively, a more familiar or non-military task could be used.

CHAPTER FIVE: CONCLUSION

The results of the study did not support the use of CLT-derived feedback adaptations in line with ERE theory in this case. High ability participants performed better when they were given more feedback, and lower ability trainees performed better when they were given less feedback. The results suggest that researchers and instructional designers should carefully consider the adaptation variable that is chosen and how adaptations will occur based on this variable. In their study, Kelley and McLaughlin (2012) suggested that trainees of higher ability should be given less feedback or support during training, this study did not support that finding. The study did not find a difference between the Hybrid and ATI adaptation approaches for high ability participants, but the direction of the results was in the hypothesized direction. I believe this comparison warrants further investigation. Very few studies to date have compared different approaches to adaptive training, and while both approaches lead to pre-post gains and higher scores on the post-test, there was no difference detected between them. The two methods could not be compared for lower ability participants, and whether the Hybrid approach is more effective for these trainees is still an empirical question. Clearly, the question of the optimal combination of adaptation variables and approaches has yet to be answered and warrants further research.

The results of the study have practical and theoretical implications. Firstly, the results suggest that CLT, and specifically the theory or ERE is not always applicable for adaptive training decisions. To illustrate, while Kelley and McLaughlin (2012) found ERE to be a useful theoretical basis for choosing feedback adaptations for a simple rule-based task with low intrinsic load, it was not supported for the current task. The task used in this study had a higher intrinsic load, demonstrating that it may only be a useful basis for simpler tasks. This may be especially true if the trainees are novices. The practical implication of this is that instructional designers should be wary when applying ERE during complex tasks. While a previous study using the same task showed ERE-derived principles could be used to adapt based on performance, the current results illustrated that it was not useful to adapt in this fashion based on ability. Researchers should carefully consider the relationship between their adaptive variables and the complexity of the task that is being adapted. In addition to examining the use of CLT and specifically ERE, this study was one of the first to compare Hybrid adaptive training and the ATI approach. While the results were not significant for higher ability participants, the trends suggested further research is needed, particularly for participants of lower ability.

APPENDIX A: DEMOGRAPHIC QUESTIONNAIRE

Demograpic Questionnaire

How old are you? __________

Gender (circle one): Male Female

What is your highest level of education (circle one)?

High School Diploma Some College Associate's Degree Bachelor's Degree

Some Graduate School Master's Degree Doctoral Degree

What was your major/focus area? ____

Are you left or right handed (circle one)? Left Right I use both equally

How often do you work with personal computers?
_____ I've never worked with a personal computer
_____ Only a couple of times ever in my life
_____ Several times a year
_____ Several times a month
_____ Several times a week
_____ At least once a day, everyday
_____ For several hours every day (over 4 hours a day)

Rate your experience with personal computers:
_____ Little or none
_____ Know a little; know Internet access, know some word processing and other software (e.g., Microsoft Word and Microsoft PowerPoint).
_____ Know quite a bit; know Internet access, know word processing well, used other software packages (e.g., Microsoft Access, FTP, SPSS, Photo Shop, etc.), and/or have done some programming (e.g., HTML).
_____ Expert; know Internet access, word processing, other software, and have much experience with different programming languages (e.g., Flash, VB, C, and Java).

Do you currently or have you previously served in the military? YES NO

If yes, what is your current status? ACTIVE RESERVIST DISCHARGED
Rating ____________ Rate___________ Rank__________________________________

Have you had any periscope related experience? Yes No
If yes, please explain the type of experience.

____

How many hours per week do you play video games? _____________

Demographic Experience Questionnaire

Please rate your experience with the following activities (circle one):

1. Playing first- person perspective video games (such as Call of Duty or Halo)

Not at all Experienced		Somewhat Experienced		Very Experienced
1	2	3	4	5

2. Playing third-person perspective or overview video games (such as Assasins Creed, God of War, or Mario Brothers)

Not at all Experienced		Somewhat Experienced		Very Experienced
1	2	3	4	5

3. Doing sculpture, painting, drawing, or other visual arts

Not at all Experienced		Somewhat Experienced		Very Experienced
1	2	3	4	5

4. Constructing verbal arguments (such as debating or writing)

Not at all Experienced		Somewhat Experienced		Very Experienced
1	2	3	4	5

5. Solving word puzzles (such as crosswords)

Not at all Experienced		Somewhat Experienced		Very Experienced
1	2	3	4	5

6. Solving picture puzzles (such as hidden picture or jigsaw puzzles)

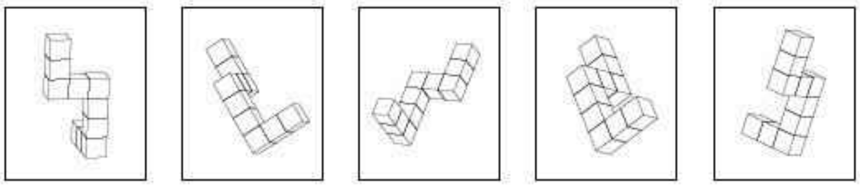
Not at all Experienced		Somewhat Experienced		Very Experienced
1	2	3	4	5

APPENDIX B: MENTAL ROTATION TEST

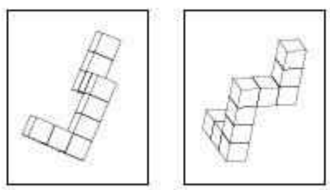
MENTAL ROTATIONS TEST (MRT-A)

This test is composed of the figures provided by Shepard and Metzler (1978), and is, essentially, an Autocad-redrawn version of the Vandenberg & Kuse MRT test.

Please look at these five figures



Note that these are all pictures of the same object which is shown from different angles. Try to imagine moving the object (or yourself with respect to the object), as you look from one drawing to the next.

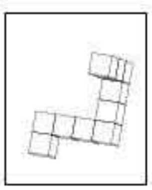


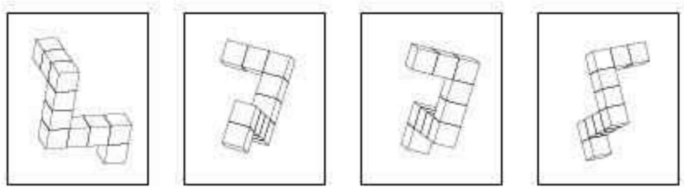
Here are two drawings of a new figure that is different from the one shown in the first 5 drawings. Satisfy yourself that these two drawings show an object that is different and cannot be "rotated" to be identical with the object shown in the first five drawings.

Now look at this object:

Two of these four drawings show the same object. Can you find those two? Put a big X across them.

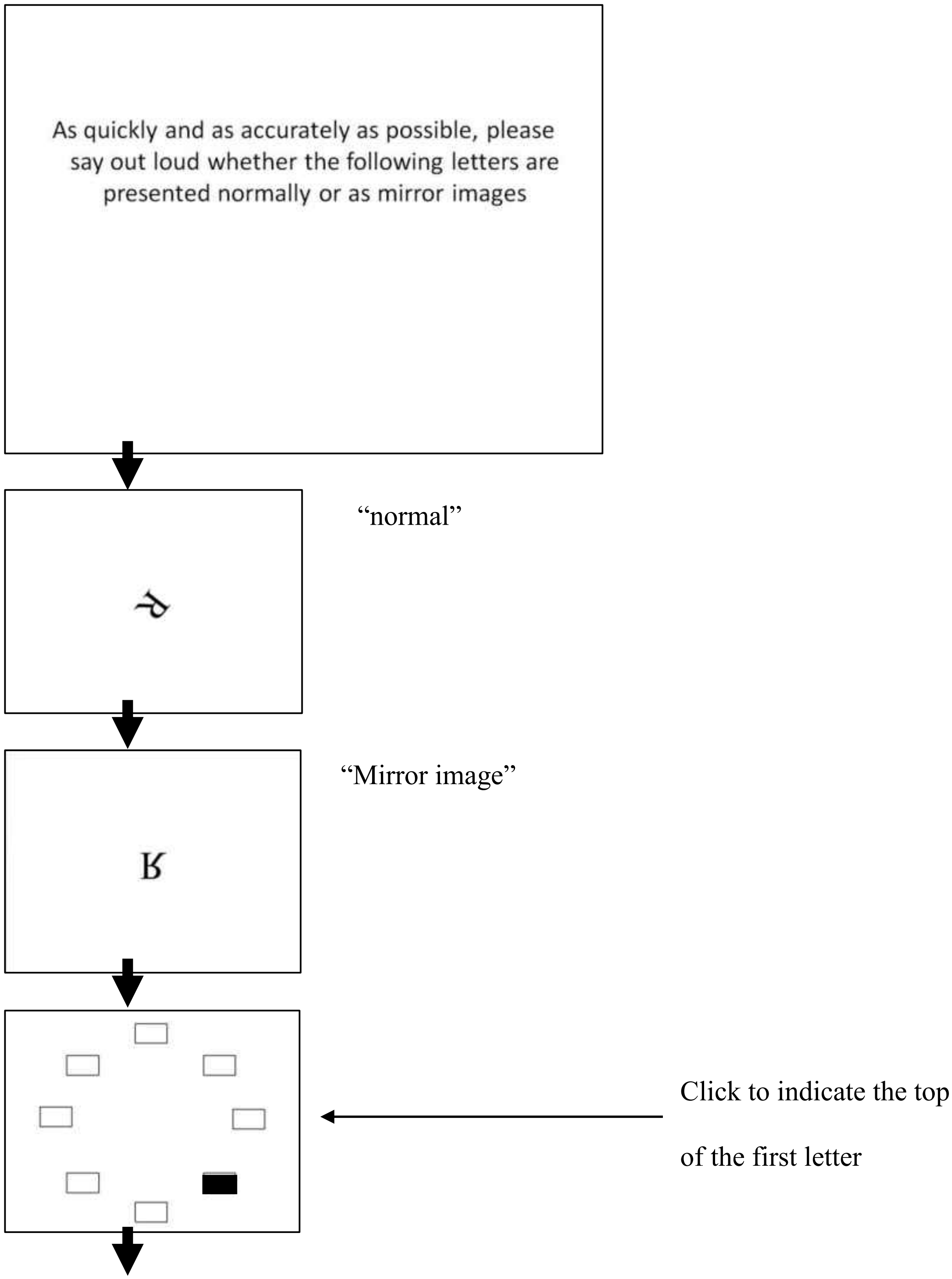
1.

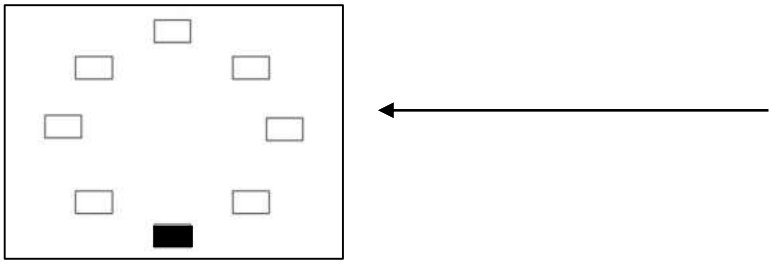




If you marked the first and third drawings, you made the correct choice.

APPENDIX C: SPATIAL SPAN TASK


As quickly and as accurately as possible, please say out loud whether the following letters are presented normally or as mirror images
“normal”
“Mirror image”
Click to indicate the top of the first letter



Click to indicate the top

of the second letter

APPENDIX D: KNOWLEDGE CHECK

Knowledge Quiz

1. The left side of the ship is called _______, the right side is called ________.
    a. Bow; Stern
    b. Port; Starboard
    c. Stern; Starboard
    d. Starboard; Port

2. The front of the ship is called the ________, the back is called the ________.
    a. Bow; Stern
    b. Port; Starboard
    c. Stern; Bow
    d. Starboard; Port

3. Angle on the Bow (AOB) can be any angle from ______-______ degrees
    a. 0-360
    b. 0-135
    c. 0-90
    d. 0-180

4. If a contact is headed directly away from the periscope operator, its AOB is
    a. 0°
    b. 45°
    c. 90°
    d. 180°

5. In the picture below, the AOB can best be described as being closest to
    a. Starboard 90°
    b. Port 90°
    c. Starboard 180°
    d. 90°

6. The front of the ship is usually _______
    a. Big
    b. Pointed
    c. Flat
    d. Red

7. One cue that you can use to determine a contact's direction of orientation is foreshortening. This means that a contact will appear to be _________ when it is not viewed at 90°
    a. Bigger
    b. Longer
    c. More distorted
    d. Farther

8. Masts or large structures on a ship will appear to get ________ when the ship is viewed at an angle smaller or larger than 90°
    a. Closer together
    b. Farther apart
    c. Taller
    d. Shorter

9. When a ship is not perpendicular to the periscope operator's line of sight, some objects may ______ others.
    a. Cover
    b. Distort
    c. Reveal
    d. None of the above

10. One way to determine the orientation of a contact is to remember that objects that are ______ appear smaller.
    a. Closer
    b. Farther away
    c. Neither
    d. Both

11. If a contact's bow appears larger than its stern, the contact has an AOB _______ 90°.
    a. Larger than
    b. Smaller than
    c. Equal to

12. What color light can be seen on a ship's port side at night or in heavy fog?
    a. Blue
    b. Green
    c. Yellow

d. Red

13. In the picture below, the contact's wake lets us know that it is moving ______ the periscope operator
    a. Towards
    b. Away from
    c. Perpendicular to

APPENDIX E: 9-POINT MENTAL EFFORT RATING SCALE

Mental Effort Rating Scale

Please indicate on the scale your level of mental effort on the task you just performed. Think only about your level of effort on the task you performed immediately preceding this questionnaire and put an X through your answer.

Very, very low								Very, very high
1	2	3	4	5	6	7	8	9

APPENDIX F: NASA TLX

NASA Task Load Index

Place a mark on each scale that represents the magnitude of each factor in the task you just performed. Indicate your answer by typing an X in the appropriate spot on each line.

Mental demand Low |____| High

Temporal demand Low |____| High

Performance Good |____| Poor

Effort Low |____| High

Frustration level Low |____| High

Rating-scale descriptions for your reference:

Title	Endpoints	Descriptions
Mental Demand	Low, High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Temporal Demand	Low, High	How much time pressure did you feel due to the rate or pace at which the task elements occurred? Was the place slow and leisurely or rapid and frantic?
Performance	Good, Poor	How successful do you think you were in accomplishing the goals of the task? How satisfied were you with your performance in accomplishing these goals?
Effort	Low, High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration level	Low, High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

APPENDIX G: SELF EFFICACY MEASURE

Self-Efficacy Questionnaire

Please indicate on the scale from 1-5 your agreement with the following statements (circle your answer). Ratings are from 1, "not at all", to 5, "very"		Not at all confident				Very confident
1.	I'm confident I can accurately find the AOB of a contact	**1**	**2**	**3**	**4**	**5**
2.	I'm confident I can quickly find the AOB of a contact	**1**	**2**	**3**	**4**	**5**
3.	I'm confident I can find the AOB of a contact at night	**1**	**2**	**3**	**4**	**5**
4.	I'm confident I can find the AOB of a contact during high sea states (waves)	**1**	**2**	**3**	**4**	**5**

APPENDIX H: DEBRIEF FORM

Debrief

Thank you for participating in today's experiment. You have participated in a study where participants play scenarios on a periscope trainer and receive different amounts of feedback (High or Low) and different types of adaptive training (Aptitude-treatment interaction training or Hybrid adaptive training). Training is a crucial component in the military, particularly in the Angle on the Bow task, because this task allows periscope officers to determine if contacts are a collision threat to their boat. The purpose of the current study is to find out which combination of instructional methods are best for improving performance. I will use your data on the task to determine which adaptive training technique works best, and how much feedback should be provided to trainees during training. I am evaluating the instruction. I am not evaluating you.

If you are interested in more information about this project, I will be happy to provide you with an abbreviated abstract of the results once the data collection is finished. To let me know if you want to receive an abstract or if you have any other questions or comments, please contact:

Carla Landsberg
Research Psychologist
(407) 380-4331
Carla.landsberg@navy.mil

Thank you for your time!

APPENDIX I: PSYCHOLOGY RESEARCH EXPERIENCE EVALUATION FORM FOR PARTICIPANTS

Psychology Research Experience Evaluation Form for Participants

Please complete this form to evaluate your experience as a participant in ________________________
Study conducted by ___________________ (Researcher)

Your Current Psychology Course(s):____

Today's Date: _______________________________

This is important to our educational efforts and the feedback you provide will aid in the evaluation and possible modification of the research participation experience. Your answers are anonymous. When you have completed this form, return it to the Psychology Department Main Office (Psychology Building – 3rd Floor).

For each question, please circle the statement that best indicates your response.

Do you clearly understand the purpose of this study?

The researcher did not explain the purpose. I did not receive a written or oral explanation of the study.	The researcher explained the purpose or gave me a written explanation of the study, but did not give me a way to ask further questions.	The researcher explained the purpose, gave me a chance to ask questions, and answered the questions I had.	The researcher explained the purpose, gave me a chance to ask questions, and answered the questions I had, and made sure I understood the purpose and implications of the study.

Was participating in this study a learning experience for you?

I completed the study, but did not receive any additional information.	I furthered my learning about the research process (informed consent, debriefing, etc.) OR this specific study (not both).	I gained information about the research process and this specific study.	I gained information about the research process, this specific study, and research that supports this study.

Were you treated with courtesy and respect?

The researcher **did not** treat me with courtesy and respect.	The researcher treated me with **some** courtesy and respect.	The researcher treated me with an **acceptable level** of courtesy and respect.	The researcher treated me with a **great deal** of courtesy and respect.

Additional comments (continue on back if necessary):

APPENDIX J: CPHS APPROVAL LETTER

DEPARTMENT OF THE NAVY
NAVAL AIR WARFARE CENTER TRAINING SYSTEMS DIVISION
12350 RESEARCH PARKWAY
ORLANDO, FLORIDA 32826-3275

IN REPLY REFER TO:
3900
Ser HRPP/214AM01
JUN 0 2 2014

From: Commanding Officer, Naval Air Warfare Center Training Systems Division
To: Carla Landsberg

Subj: APPROVAL CERTIFICATION FOR AMENDMENT 01 TO PROTOCOL NAWCTSD.2014.0001-TSD 214, TAILORING INSTRUCTION TO THE INDIVIDUAL: INVESTIGATING THE UTILITY OF TRAINEE APTITUDES FOR USE IN ADAPTIVE TRAINING

Ref: (a) 32 Code of Federal Regulation 219

Encl: (1) Protocol Information Document and Privacy Act Statement

1. Your request for Amendment 01 (AM01) to subject protocol is approved under the authority of Department of Defense Navy Assurance Number DoD N-40037, and in accordance with reference (a).

2. On 19 May 2014, the Naval Air Warfare Center Training Systems Division (NAWCTSD) Committee for the Protection of Human Subjects (CPHS) Chair conducted a review of your request for AM01. The CPHS Chair determined that your study remained minimal risk research, and qualified for an Exempt Review under Category 1. The CPHS Chair recommends approval of AM01 to TSD 214, as written.

3. Your protocol expires on 5 February 2015, or 364 days from the initial CPHS review date. An amendment does not change the protocol expiration date. Should there be a need to extend this protocol beyond the one-year approval period, as Principal Investigator, you are responsible for submitting an Application for Continuing Review (CR) and supporting documentation to the NAWCTSD CPHS in sufficient time to allow for appropriate review and re-approval before the end of the current approval period. Human subject research shall not be conducted outside of an approval period. If the approval period expires, Federal law requires that the CPHS halt the study and that work involving human subjects temporarily cease until re-approval is obtained.

4. It is your responsibility to ensure that the protocol is being followed as planned and that any adverse events or unanticipated

Subj: APPROVAL CERTIFICATION FOR AMENDMENT 01 TO PROTOCOL NAWCTSD.2014.0001-TSD 214, TAILORING INSTRUCTION TO THE INDIVIDUAL: INVESTIGATING THE UTILITY OF TRAINEE APTITUDES FOR USE IN ADAPTIVE TRAINING

problems are reported to the CPHS. Should there be a need for additional modifications to the approved protocol, you will be required to submit an amendment to the CPHS for review and approval prior to implementation of any modification to the study.

5. You are required to use the approved and date stamped Protocol Information Document, enclosure (1), when administering informed consent to potential research volunteers. Additionally, you are also required to provide a copy of enclosure (1) to each voluntary subject prior to the start of their participation in the study.

6. Should you have any questions, please contact the NAWCTSD CPHS Administrator at (407) 380-4320.

S. D. NAKAGAWA

APPROVED JUN 0 2 2014 **. This document may NOT be used after** FEB 0 5 2015 **.**

APPENDIX A.
PROTOCOL INFORMTION DOCUMENT

1. You are being asked to voluntarily participate in a research study entitled, Tailoring Instruction to the Individual: Investigating the Utility of Trainee Aptitudes for Use in Adaptive Training. You will be asked to complete training on a periscope operation task using PC-based instruction. Specifically, you will be asked to go through a series of modules that will teach you how to effectively determine a ship's angle on the bow. At the beginning of the experiment you will be asked to fill out forms asking about computer and video game experience, and measures that assess skills relevant to the task. You will then become familiarized with the task and the trainer interface and will perform a series of modules. Your performance on the task will be objectively and internally recorded by the simulator. You must be at least 18 years old to participate in this study. The researchers expect that there will be approximately 100 research subjects participating in this study. It is expected that this study will take 2.5 hours to complete. Breaks will be scheduled during this study; however, you may take breaks as needed.

2. The investigators believe that the risks or discomforts to you are as follows:

a. There are no known or expected risks to a pregnant woman, the embryo or fetus.

b. This study involves the use of typical video and/or computer games. If you have a history of seizures YOU WILL BE UNABLE TO PARTICIPATE in this study.

3. Benefits for Participation. Please understand that for this study, you will receive no direct benefit other than the knowledge that participation in this study will aid efforts to improve the performance, safety, and/or the effectiveness of U.S. Navy.

4. Compensation for Participation. For this study, the compensation you will receive will be $10.00 per hour for a maximum of three hours participation.

5. Your confidentiality during the study will be ensured by assigning you a coded identification number. Your name will not be directly

associated with any data. Any subject identifica-tion keys will be destroyed at the end of the study. This procedure will insure that your personal data cannot be used in any way that might impact your career, academic progress, or standing in your respective professional or educational communities. Please understand that all personal data or information (such as demographic data/ performance data) will be secured under lock and key until destroyed in accordance with SECNAV M-5210.1 and as required by 32 CFR 219.111(a)(7). Any data collected for this study and used across studies within NAWCTSD, will be de-identified data only.

6. Should you have questions concerning the research described in this Protocol Information Document or questions concerning research-related injury, please contact the Principal Investigator listed below. Additionally, if you so desire, you may contact the Principal Investigator for a copy of any publication resulting from this study. If you have questions concerning your rights as a research subject, please contact the CPHS Chair (see contact info provided below).

a. Principal Investigator: Carla Landsberg
   Activity: NAWCTSD
   Mailing Address: 12350 Research Pkwy, Orlando FL 32826
   Code: 4.6.5.1 Phone: 407-380-4331
   E-mail: Carla.landsberg@navy.mil

b. Project Manager: CDR Hank Phillips
   Activity: NAWCTSD
   Mailing Address: 12350 Research Pkwy, Orlando FL 32826
   Code: 4.6T Phone: 407-280-4243
   E-mail: henry.phillips@navy.mil

c. CPHS Chairman: Dr. Randolph Astwood
   Activity: NAWCTSD
   Mailing Address: 12350 Research Pkwy, Orlando FL 32826
   Code: 4.6.5.1 Phone: 407-380-4883
   E-mail: Randy.astwood@navy.mil

7. Participation in this research study is voluntary. You may choose not to participate. If you decide to withdraw from further participation in this study, there will be no penalties. To ensure your safe and orderly withdrawal from the study, please inform the Principal Investigator listed in paragraph 6.a. of this document.

APPROVED JUN 0 2 2014 **. This document may NOT be used after** FEB 0 5 2015 **.**

8. Your participation in this study may be stopped by the investigator at any time without your consent if it is believed the decision is in your best interest. There will be no penalty or loss of benefits to which you are otherwise entitled at the time your participation is stopped.

9. No out of pocket costs to you may result from your voluntary participation in this study.

10. Official government agencies may have a need to inspect the research records from this study, including your records, in order to fulfill their responsibilities. Additionally, de-identified data collected for this study may be used across studies within NAWCTSD.

11. A Privacy Act Statement concerning this research protocol is attached as Appendix B and will be provided to you prior to your participation in this study.

12. It is important that you understand what has been explained in this Protocol Information Document about your participation in this study. If you have any questions or concerns about this study and its related procedures and any risks that may be associated with your participation, please talk with the Principal Investigator listed in paragraph 6.a. All of your questions should be answered to your satisfaction and you are to receive a copy of this document for your records.

APPENDIX B.
PRIVACY ACT STATEMENT

1. Authority. 5 U.S.C. 301

2. Purpose. Performance data from computer-based training information will be collected in an experimental research project entitled: Tailoring Instruction to the Individual: Investigating the Utility of Trainee Aptitudes for Use in Adaptive Training, to assess the effectiveness of different types of instruction.

3. Routine Uses. The data collected will be used for analyses and reports by the Departments of the Navy and Defense, other U.S. Government agencies, and authorized government contractors. De-identified data collected for this study may be used across studies within NAWCTSD. Additional use of the information may be granted to non-Government agencies or individuals by the Navy Surgeon General following the provisions of the Freedom of Information Act or contracts and agreements. I voluntarily agree to its disclosure to the agencies or individuals identified above, and I have been informed that failure to agree to this disclosure may make the research less useful.

4. Voluntary Disclosure. Provision of information is voluntary. Failure to provide the requested information may result in failure to be accepted as a research volunteer in this study.

5. Appendix A. Appendix A of this protocol package contains the Protocol Information Document for this study. A copy of the approved/dated Protocol Information Document is to be provided to each research volunteer prior to the start of their participation in this study.

APPENDIX K: IRB APPROVAL LETTER

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 Exempt Human Research

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

To: **Carla Landsberg**

Date: **April 30, 2014**

Dear Researcher:

On 4/30/2014, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Project Title:	Tailoring Instruction to the Individual: Investigating the Utility of Trainee Aptitudes for Use in Adaptive Training
Investigator:	Carla Landsberg
IRB Number:	SBE-14-10256
Funding Agency:	Naval Innovative Science and Engineering (NISE)
Grant Title:	
Research ID:	N/A

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

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

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 04/30/2014 04:31:20 PM EDT

Joanne Muratori

IRB Coordinator

Page 1 of 1

APPENDIX L: TABLE OF MEANS AND STANDARD DEVIATIONS FOR ALL GROUPS

Condition	Name	Post-test median*	Post-test time (seconds)	Transfer median*	Transfer time (seconds)	Pre to post gain scores	IE outcomes	IE learning
1	Hybrid Matched High	17.73 (6.01)	13.60 (3.12)	25.91 (8.96)	14.11 (5.12)	.42 (.14)	-.61 (.78)	-.28 (.57)
2	Hybrid Matched Low	29.25 (9.64)	10.33 (2.75)	35.20 (12.16)	10.73 (4.05)	.34 (.19)	-.07 (.77)	-.28 (.67)
3	Hybrid Mismatched High	19.79 (4.96)	13.01 (3.27)	27.76 (23.49)	13.46 (3.16)	.42 (.14)	-.32 (.67)	-.05 (.69)
4	Hybrid Mismatched Low	29.08 (17.39)	12.25 (3.17)	29.27 (16.86)	12.05 (3.09)	.27 (.26)	.58 (1.38)	.27 (1.57)
5	ATI Matched High	27.90 (21.42)	12.04 (4.76)	32.00 (21.34)	13.97 (11.39)	.35 (.21)	-.12 (1.32)	-.29 (.98)
7	ATI Matched Low	17.82 (6.98)	12.30 (3.58)	19.72 (6.90)	12.76 (3.54)	.35 (.26)	-.23 (.68)	-.10 (.70)
6	ATI Mismatched High	35.96 (21.55)	12.81 (4.58)	38.54 (27.29)	12.72 (3.60)	.17 (.51)	.56 (1.26)	.35 (1.45)
8	ATI Mismatched Low	24.38 (9.02)	12.25 (4.88)	27.50 (7.80)	11.32 (2.37)	.35 (.22)	.06 (.81)	.25 (1.01)

APPENDIX M: SUPPLEMENTAL ANALYSES

Because data collected from individuals over the course of multiple trials will likely be correlated and violate the non-independence assumption of many statistical tests, including ANOVA, random coefficients growth modeling was used to analyze the during training data (Bliese & Ployhart, 2002). According to Bliese and Ployhart (2002), growth modeling can be used to examine how individuals change over time, and what the differences are in the patterns of change. A growth model analysis was performed in R to assess the effects of the three independent variables (Spatial ability, Feedback Type (Mismatch vs. Match), and Approach (ATI vs. Hybrid)) during the six sets (coded as Times 1-6) of training. The independent variables were the between subjects part of the model and Time (i.e Set) was the within subjects variable. In the first model, the within subjects model was tested using the equation:

$$\text{Training performance} = \beta_{0j} + \beta_{1j}(Time) + e_{ij} \qquad (4)$$

The between subjects model was tested using the equation:

$$\beta_0 = \gamma_{00} + \gamma_{01}(Gender) + \gamma_{02}(Pretest\ median) + \gamma_{03}(Spatial\ ability) + \gamma_{04}(Match/mismatch) + \gamma_{05}(Approach) + u_{oj} \qquad (5)$$

In equation 4, the beta weights represent the intercept and the slope of the time variable. In equation 5, the gammas represent the intercept, and the slopes of the independent variables Gender, Pretest median score, Spatial ability, Feedback match/mismatch and Approach.

First, a linear model was estimated. The results can be seen in Table 30

Table 32. Results of the linear random coefficients model analysis

Fixed Effects	Coefficient	SE	*df*	*T*
Intercept	35.96	4.72	459	7.62***
Time	-.62	.34	459	-1.82
Gender	-5.43	2.61	86	-2.07*
Pre Median	.00004	.0002	86	.22
Spatial ability	-10.46	2.61	86	-4.00***
Match/Mismatch	7.51	2.53	86	2.96**
Approach	-1.67	2.55	86	-.66

Note. ***$p < .001$ ** $p < .01$ *$p < .05$. All tests 2-tailed.

In the linear model, the within-subjects variable, Time (i.e. Set), was only marginally significant (p=.07) indicating that there was no significant growth found in this model during training. There were significant main effects of the within subjects variables Spatial ability and Match/mismatch, but no significant main effect of Approach.

Next, a quadratic model was tested. The results of this analysis are shown in Table 31.

Table 33. Results of the quadratic random coefficients model analysis

Fixed Effects	Coefficient	SE	*df*	*T*
Intercept	34.07	4.66	458	7.30***
Time 2	-30.15	12.60	458	-2.39*
Gender	-5.26	2.65	86	-1.99*
Pre Median	.00001	.0002	86	.06
Spatial ability	-10.46	2.65	86	-3.89***
Match/Mismatch	7.57	2.56	86	2.95**
Approach	-1.33	2.59	86	-.52

Note. ***$p < .001$ ** $p < .01$ *$p < .05$. All tests 2-tailed.

In this model, the time variable was significant, indicating there was significant growth during training. Similar to the linear model, there were significant main effects of the within-subjects variables Spatial ability and Match/Mismatch. No significant main effect was found for AT Approach. The effects of the non-linear analysis suggest that participants' change over time during the course of training followed a non-linear pattern. Therefore, participants' improvement during training could not be adequately detected using linear based analyses such as ANOVA; however, the non-linear random coefficients model showed that participants' performance got slightly worse early in training, and improved after the third set.

LIST OF REFERENCES

Akhras, F. N. & Self, J. A. (2000). System intelligence in constructivist learning. International *Journal of Artificial Intelligence in Education, 11*, 344-376.

Alyman, C. & Peters, M. (1993). Performance of male and female children, adolescents, and adults on spatial tasks that involve everyday objects and settings. *Canadian Journal of Psychology, 47*, 730-747.

Angeli, C., Valanides, N., & Kirschner, P. (2009). Field dependence-independence and instructional design effects on learners' performance with a computer-modeling tool. *Computers in Human Behavior,* 25, 1355–1366.

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

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 4*(11), 417-423.

Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (pp. 47-90). New York: Academic Press.

Baenninger, M. & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20*(5/6), 327-343.

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

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Barrick, M. R. & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26

Batka, J. A. & Peterson, S. A. (2005). The effects of individual differences in working memory on multimedia learning. *Proceedings of the Human Factors and Ergonomic Society*, 49, 1256-1260.

Bauer, K. N., Brusso, R. C., & Orvis, K. A. (2012). Using adaptive difficulty to optimize videogame-based training performance: The moderating role of personality. *Military Psychology, 24*(2), 148-165.

Bedwell, W. L. & Salas, E. (2010). Computer-based training: Capitalizing on lessons learned. *International Journal of Training and Development, 14*(3), 239-249.

Bell, B. S. & Kozlowski, S. W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology, 93*, 296-316.

Berliner, D. C. & Cahen, L. S. (1973). Trait-treatment interaction and learning. *Review of Research in Education, 1*, 58-94.

Bliese, P. D. & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods, 5*(4), 362-387.

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.

Brown, K. G. & Ford, J. K. (2002). Using computer technology in training: Building an infrastructure for active learning. In K. Kraiger (Ed.), *Creating, implementing, and maintaining effective training and development: State-of-the-art lessons for practice* (pp.192-233). San Francisco, CA: Jossey-Bass.

Buff, W. L., & Campbell, G. E. (2002). What to do or what not to do?: Identifying the content of effective feedback. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 2074–2078). Santa Monica, CA: HFES

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies.* New York: Cambridge University Press.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, P., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19(2),* 237-248.

Corbalan, G., Kester, L., van Merriënboer, J.J.G. (2008). Selecting learning tasks: Effects of adaptation and shared control on learning efficacy and task involvement. *Contemporary Educational Psychologist,* 33(2008), 733-756.

Corbett, A.T. (2001). Cognitive computer tutors: Solving the two-sigma problem. In M.Bauer, P.J. Gmytrasiewicz, and J. Vassileva (Eds.), User Modeling: Proceedings of the Eighth International Conference, (pp. 137-147). Heidelberg, Germany: Springer.

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 (pp 849-870). Retrieved from http://act-r.psy.cmu.edu/papers/173/Chapter_37_Intelligent_Tutoring_Systems.pdf

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671-684.

Cronbach, L. J. & Snow, R. E. (1977). *Aptitudes and instructional methods. A handbook for research on interactions*. New York: Irvington

Cronbach, L. J. & Snow, R. E. (1969). *Individual differences in learning ability as a function of instructional variables*. Final Report, School of Education, Stanford University, Contract # OEC-4-6-061269-1217, U.S. Office of Education.

Davis, J. K. (1991). Educational implications of field dependence/independence. *In* S. Wapner & J. Demick (Eds.), *Field dependence/independence: Cognitive style across the life span* (pp. 149–76). Mahwah, NJ: Lawrence Erlbaum Associates.

De Lisi, R. & Cammarano, D. M. (1996). Computer experience and gender differences in undergraduate mental rotation performance. *Computers in Human Behavior, 12*(3), 351-361.

DeLeeuw, K. E. & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*(1), 223-234.

Deutsch, T., & Tobias, S. (1980). *Prior achievement, anxiety, and instructional method.* Paper presented at the annual meeting of the American Psychological Association, Montreal, Canada.

Du, H. S. & Wagner, C. (2005). Learning with weblogs: An empirical investigation. *Proceedings of the 38th Hawaii International Conference on System Sciences, 2005*(1), 7b. doi: 10.1109/HICSS.2005.387

Gawron, V.J. (2000). *Human performance handbook.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Earley, P.C., Northcraft, G.B., Lee, C., & Lituchy, T.R. (1990). Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal, 33* (1), 87-105.

Erkstrom, R. B., French, J. W., & Harmon, H. H. (1976). *Manual for kit of factor- referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Field, A. (2013). Discovering statistics using IBM SPSS statistics 4th edition. Los Angeles: Sage.

Gawron, V. J. (2000). *Human Performance Measures Handbook*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Glaser, R., & Nitko, A. J. (1971). Measurement in learning and instruction. In R. L. Thorudike (Ed.), *Educational Measurement* (2nd ed., pp. 625–670). Washington, DC: American Council of Education.

Graesser, A. C., Conley, M. W., & Olney, A. M. (2012). Intelligent tutoring systems. In S. Graham & K. Harris (Eds.), *APA handbook of educational psychology*. Washington, DC: American Psychological Association.

Guilford, J. P. & Zimmerman, W. S. (1948). The Guilford-Zimmerman aptitude survey. *Journal of Applied Psychology, 32,* 24-34.

Hart, S. G. & Staveland, L. E. (1987). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), Human mental workload. Amsterdam: Elsevier.

Hayes, A. F. (2009). Beyond Barron and Kenny: Statistical mediation analysis in the new millennium. *Communication monographs, 76*, 408-420.

Herold, D. M., Davis, W., Fedor, D. B., & Parsons C. K. (2002). Dispositional influences on transfer of learning in multistage training programs. *Personnel Psychology*, 55, 851-869.

Horak, V. M., & Horak, W. J. (1982). The influence of student locus of control and teaching method on mathematics achievement. *Journal of Experimental Education*, 51, 18–21.

Ilgen, D.R., Fisher, C.D., & Taylor, M.T. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*, 349-371.

Just, M. A. & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92,* 137-172.

Kalyuga, S. (2005). Prior knowledge principle in multimedia learning In R.Mayer (ed.), *The Cambridge Handbook of Multimedia Learning,* (pp. 325-338). New York: Cambridge University Press.

Kelley, C. M. & McLaughlin, A. C. (2012). Individual differences in the benefits of feedback for learning. *Human Factors, 54*(1), 26-35.

Kelley, C. R. (1969). What is adaptive training? *Human Factors, 11*(6), 547-556.

Kim, P. (2010). *Measuring the effectiveness of information security training: A comparative analysis of computer-based training and instructor-based training* (Doctoral dissertation, Robert Morris University). Retrieved from http://ezproxy.lib.ucf.edu/login?url=http://search.proquest.com.ezproxy.lib.ucf.edu/docview/759244312?accountid=10003

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.

Kozhevnikov, M. & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition, 29*(5), 745-756.

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology, 20*, 397-417.

Kozlowski, S. W. J., Toney, R. J., Mullins, M. E., Weissbein, D. A., Brown, K. G., & Bell, B. S. (2001). Developing adaptability: A theory for the design of integrated-embedded training systems. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research* (Vol.1, pp. 59-123). Amsterdam: JAI/Elsevier Science.

Kulik, J. A. (1982). Individualized systems of instruction. In H. E. Mitzel, (Ed.), *Encyclopedia of educational research* (5th ed.). New York: Macmillan.

Kulik, C. C. & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior*, 7, 75-94.

Landsberg, C. R., Astwood, R. S., Van Buskirk, W. L., Townsend, L. N., Steinhauser, N. B., & Mercado, A. M. (2012). Review of adaptive training techniques. *Military Psychology, 24*(2), 9113

Landsberg, C. R., Mercado, A. D., Van Buskirk, W. L., Lineberry, M, & Steinhauser, N. (2012). Evaluation of an adaptive training system for submarine periscope operations.

Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56, 2422-2426.

Landsberg, C. R., Van Buskirk, W. L., Astwood, R. S., Mercado, A. D., & Aakre, A. J. (2011) *Adaptive training considerations for simulation-based training systems* (Special report 2010-001). Defense Technical Information Center (DTIC). Retrieved fromhttp://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA535421&Location=U2&doc=GetTRDoc.pdf

Lee, W. W., Owens, D. L., & Benson, A. D. (2002). Design considerations for web-based learning systems. *Advances in Developing Human Resources, 44*(4), 405-423.

Linn, R. L. (1993). Final report- Raising the stakes of test administration: The impact on student performance on NAEP. U. S. Department of Education, National Center for Educational Statistics. Grant RS90159001. Washington, DC.

Linn, M. C. & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child Development, 56, 1479-1498.

Locke, E. A. (1969). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance, 3*, 157-189.

Locke, E. A. & Latham, G. P. (1990). *A Theory of Goal Setting and Task Performance*. Englewood Cliffs; NJ: Prentice-Hall, Inc.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1980). *Goal setting and task performance: 1969-1980*. Retrieved from www.dtic.mil (ADA086584).

Lohman, D. F. (1988). Spatial ability as traits, processes, and knowledge. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp.181-248). Hillsdale, NJ: Lawrence Erlbaum Associates.

Massa, L. J. & Mayer, R. E. (2006). Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style? *Learning and Individual Differences, 16*, 321-335.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies; Environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin, 86*, 889-918.

Messick, S. (1994). The matter of style: Manifestations of personality in cognition, learning and teaching. *Educational Psychologist*, 29, 121–136.

Mödritscher, F., Garcia-Barrios, V. M., & Gutl, C. (2004). The past, the present and the future of adaptive e-learning: An approach within the scope of the research project AdeLE. *Proceedings of the Interactive Computer Aided Learning Conference,* Villach, Austria.

North, D. (1989). How to get the most from safety training. *Safety and Health, 139*(1), 32-37.

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

Paas, F., Renkl, A. & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38(1), 1-4*.

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

Park. O. & Lee, J. (2003). *Adaptive instructional systems*. Handbook of Research for Educational Communications and Technology, 651-684. Retrieved from http://www.aect.org/edtech/ed1/25.pdf .

Park, O. & Tennyson, R. D. (1980). Adaptive design strategies for selecting number and presentation order of examples in coordinate concept acquisition. *Journal of Educational Psychology, 72*(3), 362-370.

Park, O. & Tennyson, R. D. (1986). Computer-based response-sensitive design strategies for selecting presentation form and sequence of examples in learning of coordinate concepts. *Journal of Educational Psychology, 78*(2), 153-158.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork R. (2009). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9*(3), 105-119.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28*, 39-58.

Pew, R. W. & Mavor A. S. (Eds). (1998). *Modeling human and organizational behavior.* Washington, D. C.: National Academy Press.

Rey, G. D. & Buchwald, F. (2011). The expertise reversal effect: Cognitive load and motivational explanations. *Journal of Experimental Psychology, 17*(1), 33-48.

Schaefer, P.S. & Dyer, J.L. (2011). Bridging the gap between adaptive training research and army practice. *Military Psychology*, 24, 194-219.

Schmidt, R. A. & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Seyler, D. L., Holton III, E. F., Bates, R. A., Burnett, M. F., & Carvalho, M. A. (1998). Factors affecting motivation to transfer training. *International Journal of Training and Development, 2*(1), 2-16.

Shah, P. & Miyake, M. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. Journal *of Experimental Psychology: General, 125*(1), 4-27.

Shearon, B. T. (2001). *The cost effectiveness of west coast distributed simulation training for the Pacific fleet* (Master's thesis). Retrieved from dtic.mil. (ADA406612)

Shin, E.C., Schallert, D.L., & Savenye, W.C. (1994). Effects of learner control, advisement, and prior knowledge on young students' learning in a hypertext environment. *Educational Technology Research and Development, 42*, 33-46.

Shute, V. (2007). *Focus on formative feedback* (RR-07-11). Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/RR-07-11.pdf

Shute, V. J., & Zapata-Rivera, D. (2008). Adaptive technologies. In J. M. Spector, D. Merrill, J. van Merriënboer, & M. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 277–294). New York, NY: Lawrence Erlbaum Associates.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*, 489-528.

Snow, E. R., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76,* 347–376.

Smither, J.W., Wohlers, A.J., & London, M. (1995). A field study of reactions to normative versus individualized upward feedback. *Group Organization Management*, 20, 61-89.

Stottler, R. H., & Vinkavich, M. (2006). *Tactical Action Officer Intelligent Tutoring System (TAO ITS).* Defense Technical Information Center (DTIC). Retrieved from http://handle.dtic.mil/100.2/ADA454597.

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

Tennyson, R. D., & Rothen, W. (1977). Pretask and on-task adaptive design strategies for selecting number of instances in concept acquisition. *Journal of Educational Psychology, 69*(5), 586–592.

Tennyson, R. D., & Christensen, D. L. (1988). MAIS: An intelligent learning system. In D. Jonassen (Ed.), *Instructional designs for micro-computer courseware*, 247-274. Hillsdale, NJ: Lawrence Erlbaum Associates.

Training (2011). *2011 training industry report.* Retrieved from http://www.trainingmag.com/sites/default/files/trg_2011IR.pdf

Tuovinen, J. E. & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*(2), 334-341.

U.S. Office of Management and Budget. (2012). Fiscal Year 12 of the U.S. Government (2012 Ed.). Washington, DC: U.S. Government Printing Office. Retrieved December 5, 2012 from the World Wide Web: http://www.whitehouse.gov/sites/default/files/omb/budget/fy2012/assets/defense.pdf.

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

Van Buskirk, W., Cornejo, J., Astwood, R., Russell, S., Dorsey, D., & Dalton, J. (2009). A theoretical framework for developing systematic instructional guidance for virtual environment training. In D. Schmorrow, J. Cohn, & D. Nicholson (Eds.), *The PSI* *handbook of virtual environments for training: Vol. 1. Learning, Requirements, and Metrics* (pp. 97–113).Westport, CT: Praeger Security International.

van Gog, T. & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16-26.

van Merriënboer, J. J. G. & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147-178.

Vandenberg, S. G. & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599-601.

Vandewaetere, M., Desmet, P, & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior, 27*, 118-130.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197-221.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? Cognitive Science, 31, 3-62.

Williams, T. C. & Zahed, H. (1996). Computer-based training versus traditional lecture: Effect on learning and retention. *Journal of Business and Psychology, 11*(2), 297-310.

Wolf, L. F. & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*(3), 227-242.

Yeany, R.H., Dost, R.J., & Mattews, R.W. (1980). The effects of diagnostic- prescriptive instruction and locus of control on the achievement and attitudes of university students. *Journal of Research in Science Teaching*, 17, 537–546.