INVESTIGATING THE ROLE OF COGNITIVE LOAD IN SYNTHETIC LEARNING ENVIRONMENTS FOR TRAINING

by

JOELENE GOH
M.S. University of Central Florida, 2018

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2018

Major Professor: Clint Bowers

ABSTRACT

Ensuring effectiveness of training programs has been a dominant theme in the training industry, and is constantly evolving with the steady incorporation of emerging technology. This field study offers an investigation into the intersection between the applied and research world, and examines the practicality of recommended best practices for implementing synthetic learning environments (SLEs) in the military. Specifically, cognitive load has been identified as a significant factor in influencing the effectiveness of training programs. Research on this topic has focused on utilizing the affordances of SLEs to decrease cognitive load imposed by the material and system in order to allow for more cognitive resources to be allocated towards schema construction and automation. Therefore, this study was derived from a need to ensure that the introduction of SLEs into training programs did not hinder learning or training transfer by comparing the performance outcome measures from two SLEs, Virtual BattleSpace 3 (VBS3) and the Military OpenSimulator Enterprise Strategy (MOSES). Based on concepts of cognitive load, it was possible that any group differences could be explained by the varying levels of cognitive load imposed by either system. Furthermore, the specific system could influence the strength of the effect of cognitive load on performance measures. A conditional process analysis model was constructed from the theorized relationships, and the bootstrap method was used to analyze the model. Research findings indicated no support, and discussions delved into possible explanations for results of the study, limitations, and recommendations for future research. While the analyses were nonsignificant, this was the first study investigating the difference between the VBS3 and MOSES platform, and indicated no difference in impacting performance.

Additionally, because MOSES is a free, open source platform, this study could support industries that are looking for cost-effective methods to expand training programs in the direction of SLEs.

To the best parents, Jeremiah and Evelyn, for your everlasting and unconditional love. It is impossible to put into words the impact you have had on my life. I am here because of you.

To my siblings, thank you for your support and a lifetime of fun.

Finally, to Dr. Richard Milich, Dr. Elizabeth Lorch, and Dr. Richard Smith:

As educators, you hold the power to influence a student's life forever,
and I was very fortunate to have had you as my advisors.

ACKNOWLEDGMENTS

To my committee chair, Dr. Clint Bowers: Thank you for your wisdom, guidance, and humor throughout this project, for making time for our many meetings, and for your belief in my abilities. To my committee members, Dr. Daniel Barber and Dr. Steve Jex: Thank you for your patience, effort, and invaluable insight. This project would not have been possible without the help of my committee.

To the Prodigy Lab at the Institute for Simulation and Training: For the opportunity to work on the MOSES project.

Dr. Barbara Truman, it has been such a pleasure working with you. Thank you for teaching me about games, plants, and everything under the sun.

TABLE OF CONTENTS

LIST OF FIGURES x

LIST OF TABLES xi

CHAPTER ONE: INTRODUCTION 1

- Background 1
- Purpose of Study 2

CHAPTER TWO: LITERATURE REVIEW 4

- Introduction 4
- Training Evaluation 4
- Training Effectiveness 7
- Training is Broke 8
- Technology in Training 11
- Cognitive Load Theory 14
- Maximizing Training for Infantry Soldiers 19
- Proposed Study 22

CHAPTER THREE: METHODOLOGY 24

- Overview 24
- Measures 26

Performance Ratings 26

Cognitive Load 27

Demographics 28

CHAPTER FOUR: FINDINGS 29

Demographics 29

Descriptive Statistics 29

Preliminary Analysis 31

Conditional Process Analysis 36

CHAPTER FIVE: DISCUSSION AND CONCLUSION 40

Summary 40

Discussion 40

Implications 58

Theoretical Implications 59

Practical Implications 60

Limitations 63

Areas of Future Research 64

Acknowledgments 65

APPENDIX A: INFORMED CONSENT FORM 66

APPENDIX B: PERFORMANCE MEASURES 71

APPENDIX C: NASA-TLX QUESTIONNAIRE 74

APPENDIX D: DEMOGRAPHICS QUESTIONNAIRE 76

APPENDIX E: UCF IRB APPROVAL LETTERS 80

APPENDIX F: ARL IRB APPROVAL LETTERS 85

REFERENCES 90

LIST OF FIGURES

Figure 1. A typical Virtual BattleSpace 3 station. 21

Figure 2. A typical MOSES station. 21

Figure 3. Conceptual conditional process analysis model. 23

Figure 4. Soldiers at the STX lanes. 26

Figure 5. Example data evaluation sheet. 27

Figure 6. Statistical conditional process analysis model. 37

LIST OF TABLES

Table 1: Demographic profile of participants 30

Table 2: Descriptive statistics for overall and average task ratings 32

Table 3: Correlation analysis of performance measures 34

Table 4: Results for the model (PROCESS Model 74, Hayes, 2017) investigating conditional process relationship between condition of training program, mental demand, and performance measure for React to Improvised Explosive Device Overall. 38

Table 5: Results for the model (PROCESS Model 74, Hayes, 2017) investigating conditional process relationship between condition of training program, mental demand, and performance measure for React to Near Ambush Overall. 39

Table 6: Results for the model (PROCESS Model 74, Hayes, 2017) investigating conditional process relationship between condition of training program, mental demand, and performance measure for React to Far Ambush Average. 39

CHAPTER ONE: INTRODUCTION

Background

Training and development programs have been well-established as a vital element for success, whether used in educational contexts, military training, or organizations (Neily, et al., 2010; Harman, et al., 2008), and have been referred to by Goldstein and Ford (2002) as "…a systematic approach to learning and development to improve individual, team, and organizational effectiveness" (as cited in Aguinis & Kraiger, 2009, p. 452). The history of training can be traced to the earliest form of on-the-job training, generally utilized in the practice of apprenticeships, and evolved alongside the Industrial Revolution. The burgeoning expansion of factories generated not only an immense increase in the demand for workers, but also facilitated the improvement of the process to rapidly train skills required for jobs as well (Sleight, 1993).

The industry surrounding training has grown spectacularly in the past few decades since, and is now widely accepted as an integral feature of the workplace. In 2017, U.S.-based corporations reportedly invested $93.6 billion on total training expenditures, a hefty $13 billion increase from the previous year, and required an average of 47.6 hours of training per employee (Industry Report, 2017). Organizations also vary in the modality of training delivery, with an increasing number venturing into the ample affordances of technology-based training. In 2017, a survey of large companies of 10,000 employees or more revealed that 39.5% of training programs were instructor-led, while 60.5% used some type of technology (virtual classroom, computer-based methods, mobile-based methods, etc.). The advent of a flurry of technological advances has resulted in organizations' currently most frequently anticipated purchases being online training tools (40%, a 5% increase from the previous year), which has been an indisputable factor in the increasing budget for training and development programs (Industry Report, 2017).

The increase in interest towards training has no doubt stemmed from research from the early 21st century investigating the benefits of training, which has resulted in resounding support for its positive effects on the individual, team, and organizational level. An overview of literature by Aguinis and Kraiger (2009) point to an increase in many desired outcomes, such as job performance, declarative and procedural knowledge, self-efficacy, and leadership skills for the individual. Teams undergoing training experienced benefits such as improved team effectiveness, which may have led to a decrease in field errors—results that are highly important for high-risk jobs, such as aviation. Organizations reported a positive relationship between training and profitability, overall effectiveness, and objective measures of performance.

Purpose of Study

The billions of dollars poured into training warrant continual extensive research for ensuring its effectiveness and desired return on investment for the organization, and as a result, "…there has been nothing less than an explosion in training-related research in the past 10 years" (Salas & Cannon-Bowers, 2001, p. 472). A multitude of studies have thus been published exploring topics such as training techniques, methods for evaluation, and the outcomes and valid criteria of interest (e.g. Colquitt, LePine, & Noe, 2000; Arthur, Bennett, Edens, & Bell, 2003; Keith & Frese, 2008; Salas, et al., 2008). The purpose for this study was to examine the role of cognitive load in successful training transfer between two types of serious game platforms using measures from a training evaluation.

CHAPTER TWO: LITERATURE REVIEW

Introduction

While the advancement of the training industry dates back to the expansion of factories during the Industrial Revolution, its popularity is likely a result of the abundant literature demonstrating its benefits and establishment as an industry. The field of training focuses on two main aspects: training evaluation and training effectiveness.

Training Evaluation

The seminal works of Kirkpatrick (1959a; 1959b; 1960a; 1960b; 1967) and his delineation of the four dimensions of evaluation of training have been referenced by an expansive audience to lay the foundation for program evaluations worldwide. Defined by Alvarez, Salas, and Garofano (2004) as “a measurement technique that examines the extent to which training programs meet the goals intended” (p. 387), training evaluations are necessary for demonstrating that successful training has occurred, and are generally of great interests of shareholders and/or owners.

Kirkpatrick’s method has been viewed as adaptable and straightforward, and the simple nature of the model is a major contribution to its ubiquity in the training industry (Alvarez, Salas, & Garofano, 2004; Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997). Level 1, “Reactions,” are used to gauge the trainees’ feelings about the program. These affective reactions and utility judgments are emotionally-guided opinions, and are commonly measured by enjoyment during training and the perceived usefulness of the program (Kirkpatrick, 1967;

Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997). Level 2, “Learning,” consists of knowledge-based and behavioral/skill-based tests, and can be administered immediately after the training program or after a certain interval of time. Level 3, “Behavior,” specifically refers to how the training is used on the job, and are skills and/or behaviors that have been applied in the workplace. A plethora of research has been conducted around training transfer, the factors influencing transfer, and the pervasive “transfer problem” that organizations face, which still remains a point of contention today (Baldwin & Ford, 1988; Ford & Weissbein, 1997). The last level, “Results,” are objective measures on the organizational level, such as profitability and customer satisfaction. This level is viewed as the “ultimate” criteria, as it is, ultimately, the desired outcome from a training program (Brogden & Taylor, 1950).

Criticisms regarding Kirkpatrick’s levels and the ways they have been interpreted and implemented have surfaced following its growing popularity. These rising contentions surround the ambiguousness that have led to assumptions which were, more than likely, unintentional of Kirkpatrick (Alliger & Janak, 1989). For example, these levels have been misconstrued in literature as each one being more informative than the previous level, a misunderstanding that may have been driven by advocates of the notion of the “dollar criterion” (Brogden & Taylor, 1950), or Level 4, being the decisive measure in evaluation. However, this is problematic as an organization may not require change on all four levels. For instance, a company that is exceeding its profit margins may desire a change in its culture, which would necessitate simply an evaluation of the reactions; evaluations of the other levels would be uninformative. This assumption may also redirect the focus of training programs and evaluations onto solely what can be quantified, disregarding other equally important consequences of training such as human resources management.

Another assumption originates from the idea that each level is causally linked, and therefore, one level cannot be preceded by the other (Alliger & Janak, 1989). For example, while it may be deduced that having positive affective feelings towards a training program is more likely to facilitate learning (Level 1 causing Level 2), and negative affective feelings would hinder learning, research on this topic has disproved this perception by showing a negative or neutral correlation between the two levels (Kaplan & Pascoe, 1977; Rodin & Rodin, 1972). As such, it would be unfitting, and perhaps even detrimental, to not only assume that causality exists between the levels, but the direction as well. Alliger and Janak (1989) also examine the belief that there is a positive correlation between each level, a point which has been rendered illogical and incorrect by discussions about the previous assumption.

Nevertheless, Kirkpatrick's model persists as a reference point for the subsequent models that have been proposed in an attempt to address criticisms. For example, Tannenbaum, Cannon-Bowers, Salas, and Mathieu (1993) included training performance and transfer performance into the model and suggested that reactions to the training was unrelated to the other levels. Holton (1996) excluded reactions from his model, asserting that reactions were not an outcome of interest of training programs, but rather acted as a mediator and/or moderator variable between motivation to learn and learning itself. An additional evaluation strategy by Kraiger (2002) refocused evaluation onto three dimensions: the training content and design of the training, reactions of the trainees, and outcomes on the organizational level.

The four levels of training evaluations remain consistently used by organizations, regardless, and rationale for the basis of how the performance measures are constructed and evaluated in this study centered upon Kirkpatrick's model, focusing on the "learning" from the training.

Training Effectiveness

Training effectiveness, conversely, differs from training evaluation in that it focuses on the context surrounding the training program, and is described to be "…the study of the individual, training, and organizational characteristics that influence the training process before, during, and after training" (Alvarez, Salas, & Garofano, 2004, p. 389). The key to training effectiveness is derived from a thorough needs analysis, from which the context of the situation can be used to diagnose the problem and prescribe an effective training program.

Individual characteristics may consist of personality traits, experience, and expectations; organizational characteristics include organizational culture and climate, policies, and history; and training characteristics encompass the design of the program, instructional style, and practice (Alvarez, Salas, & Garofano, 2004). It is crucial to analyze factors of the context of the situation, because their influence over training programs may either facilitate or hinder efforts for successful training transfer. Proposed models for training effectiveness are typically constructed based upon the relationship between the context of the situation and the variables that are targeted by training evaluation. Similar to training evaluation, there have been several models recommended for optimizing training.

In an attempt to parse out specific characteristics that influence transfer of training, Baldwin and Ford (1988) outlined input (trainee characteristics, training design, and work environment) and output (learning and retention) factors that affected the overall transfer process both directly and indirectly. Holton and Baldwin (2000) revised Baldwin and Ford's model by including specific characteristics that contribute to the output factors, such as ability and motivation of the individual, training content and design, and organizational support. Furthermore, another training effectiveness strategy proposed by Holton (1996) gained support through the development and validation of the Learning Transfer System Inventory, with the introduction of novel primary influences of ability and secondary influences of motivation and environmental elements.

To augment the implementation of training programs, Broad and Newstrom (1992), along with echoing Baldwin and Ford's (1988) summation of the relationships between the context of the situation with learning and transfer, advised tactics to enhance training effectiveness before, during, and after the training programs. Taylor, Russ-Eft, and Chan (2005) and Wexley and Baldwin (1986) also presented alternative strategies to maximize transfer through design elements such as goal setting, training of supervisors, and feedback through rewards and sanctions.

Training is Broke

There has been a wide range of studies published in support of training and its positive effects on desired outcomes. For example, in a meta-analysis by Arthur, Bennett, Edens, and Bell (2003), the authors were able to find medium to large effect sizes for organizational training effectiveness across all levels of their evaluation criteria (using Kirkpatrick's levels of training evaluation). Ashenfelter (1978), Ashenfelter and Card (1985), and Card and Sullivan (1987) reviewed the success of a training program, and found a positive effect on trainee earnings, and Burke and Day's (1986) analysis of a managerial training program revealed its moderate effectiveness. It would be erroneous to assert a paucity of literature in favor of training (Collins & Holton, 2004; Salas, Tannenbaum, Kraiger, & Smith-Jentsch, 2012).

However, further research on this topic has highlighted its spurious effect on desired outcomes and issues surrounding training. In addition to misinterpretations of Kirkpatrick's levels of training evaluation (Alliger & Janak, 1989), several limitations have been of great concern to both academic and practical fields. Surveys and studies conducted on training evaluations have shown that organizations tend to emphasize heavily on evaluating programs on the first level of Kirkpatrick's model, with decreasing percentages through the levels (ASTD, 1997, as cited in Alliger et al., 1997; Brandon Hall Group, 2015, p. 12; Mimeo, 2017). This is problematic because such affective measures have yielded an extremely low correlation with on-the-job performance (.07), which conflicts with the overall outcome goals of a training program (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997).

Moreover, uncertainty of appropriate selection of criteria to conceptualize outcomes has been another issue with training programs as it may impede accurate training evaluations and effectiveness strategies (Baldwin & Ford, 1988). For example, Ashenfelter and Card (1985) concluded that using earnings in differing years as the criteria yielded contradictory results, which may affect how training effectiveness can be managed. Arthur, Tubré, Paul, and Edens

(2003) studied the relationship between affective reactions and learning as a criterion, based on previous research demonstrating medium to large effect sizes between all levels of training and learning. The authors found a resulting small to zero relationship and surmised that the two variables assessed different aspects of teaching effectiveness (p. 282). While constant research is being conducted to determine precise relationships between training and its outcomes (e.g. Cheng & Ho, 2001; Collins & Holton, 2004; Colquitt, LePine, & Noe, 2000; Salas & Cannon-Bowers, 2001), there will always be a need to investigate and establish accurate criteria in order to improve training effectiveness.

Ensuring transfer of training has also been a predominant theme in the training industry, and has been identified and labelled as the “transfer problem” (Michalak, 1981; Baldwin & Ford, 1988). Georgenson (1982) estimated a meager 10% of transfer from a training program to on-the-job behaviors, and a survey by Saks (2002) reported a 34% retention rate after a year of participating in the training program. A particularly enlightening study by IBM (2008) illuminated this transfer conundrum further with several appalling results. Over 50% of the sample from the survey stated that their organization’s main workforce-related problem was the failure to successfully develop skills for current and future job requirements, and almost 40% of employee skills were identified as misaligned with organizational priorities. When asked how effective employment development techniques were, only 40% selected “Very effective” for on-the-job training, with substantial decreases by type of technique (p. 29). Furthermore, despite the tools and strategies that have been identified to accurately calculate return on investment, Aguinis and Kraiger (2009) mentioned a dearth of published studies indicating desired returns from training programs.

To address these issues, training experts have attempted to review significant characteristics and suggest training designs to maximize transfer. Burke and Hutchins (2007) identified factors that influenced transfer (specific factors within learner characteristics, training design and delivery, and work environment) to summarize recommendations, and Salas & Stagl (2009) emphasized a focus on systematic designing of training, delineating methodical steps to follow to ensure results. However, a different strategy to target the issue of ineffective training has emerged with the rapid advancements of technology.

Technology in Training

Serious games are defined by Zyda (2005) as "…a mental contest, played with a computer in accordance with specific rules, that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives" (p. 26), and are encompassed under synthetic learning environments (SLEs) which refer to a "subset of technology-enabled instructional systems" (Cannon-Bowers & Bowers, 2009, p. 229). Serious games have developed alongside the "gamer generation" in which increasing exposure to games has shaped the way individuals learn (Kirkley, Tomblin, & Kirkley, 2005). The term "serious game" was posited to have first been used in 2002 by David Rejeski and Ben Sawyer with the beginning of the Serious Game Initiative, and was initially contrived for training purposes for specific jobs (Bellotti, Berta, & De Gloria, 2010). However, the evolution of technology has resulted in its extensive assimilation into organizations, education, and the military (Bonk & Dennen, 2005).

Various industries have capitalized upon the affordances of serious games, and there has been a noticeable shift towards the incorporation of technology into numerous contexts (e.g. Mikropoulos & Natsis, 2011; Haque & Srinivasan, 2006; Harman, et al., 2008; Vogel, et al., 2006). A *State of the Industry Report* published in 2005 by the American Society for Training and Development reported that 28.1% of training hours were technology-based (as cited in Bell & Kozlowski, 2007), and within large companies, training methods conducted with technology have now increased to 60.5% from 33% in 2005 (Industry Report, 2005; Industry Report, 2017).

Research investigating the benefits of technology have been overwhelmingly positive. Major selling points, such as cost reduction, a decrease in training time, and broader accessibility have supported the argument for technology in training (Bell & Kozlowski, 2007). Furthermore, because training transfer is optimized when stimulus and response features are identical in both training and actual settings (Woodworth & Thorndike, 1901), technology can augment training in high-risk situations by allowing trainees to practice within a harmless and consequence-free environment while maintaining desired levels of fidelity (Farrington, 2011).

Subsequently, it is of concern that as organizations begin integrating technology into training programs, evaluating the effectiveness remains the focal point. A meta-analysis by Wouters, van Nimwegen, van Oostendorp, and van der Spek (2013) revealed that serious games were more effective for learning and retention purposes, with further gains from multiple practice sessions and team-based training. Backlund and Hendrix (2013) conducted a review of serious game literature and found an increase in knowledge, problem solving skills, and learning motivation. These encouraging results, among many others, have led to an eruption of serious games being used for training purposes.

While many advocates of serious games have touted its benefits, however, closer examination into its effects have yielded inconsistent findings. Sitzmann's (2011) meta-analysis of computer-based simulation games echoed the positive relationship between computer-based simulation games and outcomes such as retention and an increase in knowledge, but showed a moderation effect from factors such as characteristics of the game, active versus passive teaching styles, and trainee engagement. Similarly, Vogel et al. (2006) summarized cognitive benefits from technology in a meta-analysis but noted that gender and interaction with the game (whether personally or through an instructor) moderated its effects. Girard, Ecalle, and Magnan (2013) explicitly detailed the spurious nature of technology, emphasizing the inability to accurately demonstrate effectiveness of serious games for training.

In a study investigating the use of simulation for aviation training, Salas, Bowers, and Rhodenizer (1998) suggested concentrating on instructional features, stressing that there has been a lack of consideration on the learner as well as learning styles. The authors proposed that features facilitating learning and performance outcomes, such as scenario design and feedback mechanisms, should be built into the design of the simulation to ensure effectiveness of the training program. Furthermore, there is an existing assumption within the technological field that "More Is Better" with increased realism through superior graphics (Salas, Bowers, & Rhodenizer, 1998, p. 202); however, there has been no evidence of a relationship between high fidelity of a simulation and successful transfer. Training experts have cautioned against haphazardly developing and incorporating serious games and simulation for training, and assert that a thorough analysis of the training needs should be required in order to properly determine the appropriateness of technology on a case-by-case basis (Salas & Cannon-Bowers, 2001; Sitzmann, 2011).

Although technology has been proffered as a possible solution to the ambiguity of training effectiveness, individual factors influencing transfer may obstruct a favorable outcome. For instance, cognitive ability has been well-established as a relevant factor in training effectiveness, with additional research focusing on the amount of cognitive resources available and the amount required as a component of training (Kanfer & Ackerman, 1993; Bell & Kozlowski, 2007; Noe, 1986). Instructional design of the program is postulated to be a potential burden on cognitive resources, and has been a topic of interest for the training industry (Burke & Hutchins, 2007). Put plainly, "[to] be effective, serious games must incorporate sound cognitive, learning, and pedagogical principles into their design and structure" (Greitzer, Kuchar, & Huston, 2007, p. 1). It is then crucial to first be certain that the modality of training does not hinder transfer and training effectiveness.

Cognitive Load Theory

Cognitive load theory has been recognized as the "major factor that determines the success of an instructional intervention" (Paas, Tuovinen, Tabbers, & Van Gerven, 2003, p. 64), and was introduced by Sweller through his formative works (1988; 1994). He suggested that difficulties associated with learning can be manipulated through instructional design by schema acquisition and automation. Schemas are defined as "…a cognitive construct that organizes the elements of information according to the manner with which they will be dealt" (Sweller, 1994, p. 296) and can be thought of as heuristics the brain utilizes when in certain familiar conditions.

The creation of domain specific schemas can greatly reduce cognitive load by allowing individuals to classify and comprehend multiple components of information as a single component, thereby decreasing time and effort required to understand a concept (Tannenbaum & Yukl, 1992; Kalyuga, Chandler, & Sweller, 2001). This is because schemas are held in the long-term memory, which has unlimited capabilities, as opposed to the working memory, which is limited to the "magical number seven" items (Atkinson & Shiffrin, 1968; Miller, 1956). If the instructional design of the training program imposes a cognitive load that exceeds the abilities of a user's working memory, the user will not have remaining resources to comprehend and learn the training material.

Therefore, schemas are extremely beneficial for the purposes of serious games for training endeavors as expert users possessing this domain specific schema will already have a preexisting comprehension about the instructional material and will be able to process it as a single element. This will permit users to allocate more cognitive resources towards the material in the training program instead, and facilitate greater learning of the material (Sweller, 1988). For example, although users may not have had specific experiences with the technology used during a training program, their generalized knowledge of technology will help guide their usage by knowing the functions of certain buttons and keys (Paas & Van Merriënboer, 1994).

Sweller (1994) expanded upon the model of cognitive load further by identifying different sources of cognitive load. Intrinsic load is caused by the inherent complexity of the material itself (Sweller, 1994); for example, the subject of calculus is fundamentally more difficult to comprehend than algebra, and therefore an individual attempting to learn calculus will experience more intrinsic cognitive load than an individual attempting to learn algebra. Extraneous cognitive load, however, was explained to be a result of the instructional design of the program, generally due to instructional procedures having been developed without consideration to cognitive architecture of the brain (Paas, Renkl, & Sweller, 2003). The last dimension of cognitive load is germane load, which is attributable to the instructional design of the program used to teach the material (Paas, Renkl, & Sweller, 2003). While extraneous cognitive load inhibits learning and is an unwanted consequence of the instructional design, germane cognitive load assists in the creation of forming domain specific schemas.

It is known that intrinsic, extraneous, and germane cognitive load are additive and cannot be greater than the amount of cognitive resources available (Paas, Renkl, & Sweller, 2003). Because intrinsic load is innate to the material, it can only be reduced by schemas that an individual possesses and not by the design of the program. Cognitive resources that are not allocated towards comprehending the material will be distributed amongst extraneous and germane cognitive load. If extraneous load is minimized by the instructional design, this will allow a larger portion of resources towards germane cognitive load, which will enable more schema development. The next time an individual is presented with this stimulus, the established schema will decrease the overall intrinsic cognitive load, allowing more working memory capacity to process the material and continue developing more advanced schemas. As the individual repeats this cycle, they will slowly cultivate an increasingly efficient schema, which will enhance learning (Paas, Renkl, & Sweller, 2003, p. 2). Therefore, it is optimal to reduce extraneous cognitive load caused by the instructional design to improve training effectiveness.

Automatic processing is an equally important aspect of cognitive load and is defined as the "…activation of a learned sequence of elements in long-term memory that is initiated by appropriate inputs and then proceeds automatically…" (Schneider & Shiffrin, 1977, p. 1). This is an uncontrolled reaction in which activation requires no attention, and because it does not stress the capabilities of the working memory, imposes no cognitive burden (Shiffrin & Schneider, 1977). Automation allows for the individual to be efficient with restricted cognitive resources and limits of a working memory. As automatic processing occurs through permanently developed reactions to a stimulus, however, it requires extensive amounts of exposure for automation to be established.

Schemas have the ability to become an automatic process, and construction of schemas are dependent upon being able to automatize actions as well. Being able to automatically process stimuli and the reactions required will allow the working memory to devote more resources towards not only performing the tasks but also developing schemas for future interactions with the stimuli (Sweller, Van Merriënboer, & Paas, 1998; Van Merriënboer & Sweller, 2005). It is imperative when creating a training program to ensure that the design is aiding in the automaticity of tasks and construction of schemas.

Proper instructional design will minimize the split-attention effect, in which individuals are having to split their attention between integrating the material enough to where sense can be made of it, and actually comprehending the material (Chandler & Sweller, 1992). If the environment of a serious game training program is conducive to the individual’s usage, there will be less attention required to understand using the system, which will allow more attention and cognitive resources towards the actual material.

The four-component instructional design model, proposed by Van Merriënboer, Jelsma, and Paas (1992), is a framework for designing instructional programs in consideration of cognitive architecture. This model aims at strategies to develop automatic processing for tasks in order to mitigate cognitive load. Component 4, “[analysis] of supportive knowledge to perform nonrecurrent skills” (p. 27), is derived from the idea that while an individual may not be in an identical situation that he or she has faced before, preexisting schemas can allow learners to apply their knowledge to a generalized situation. For serious games, although trainees may not have had practice with the specific system used for the training program, if they have prior knowledge about the interface of games in general, they may be able to apply that knowledge to the game to facilitate in their assimilation of the information and understanding of the material. As a result, less cognitive resources will be spent towards operating the game, and more towards experiencing the training program.

The crux of the cognitive load theory issue for serious games used in training programs lies in the instructional design. Research has indicated that “…an analysis of both intrinsic and extraneous cognitive load can lead to instructional designs generating spectacular gains in learning efficiency" (Sweller & Chandler, 1994, p. 185). The value of careful consideration of all factors influencing training effectiveness when building a training program cannot be emphasized enough. Therefore, the U.S. Army Research Laboratory, capitalizing upon recent advancements in technology to optimize its training programs, has endeavored to evaluate the effectiveness of training programs conducted in recent technological innovations for future improvement (GAO, 2016).

Maximizing Training for Infantry Soldiers

Virtual Battlespace 3 (VBS3) is one of the most commonly used serious games approved for use in the Army and has served the purpose of a preliminary investigation into integrating technology into training (GAO, 2016). The military's investment is based on research showing that hands-on, cooperative play when using serious games led to experiential learning, which is more effective than traditional "live" training, such as lectures (Dickey, 2005; Jarmon, Traphagan, Mayrath, & Trivedi, 2009). As with several benefits posited by the literature, serious games provide a cost-effective and consequence-free environment for soldiers to practice and develop skills necessary for high-stakes, unpredictable situations that are faced daily in the field (Munro, Patrey, Biddle, & Carroll, 2015).

However, the initial studies exploring VBS3 have elicited questions surrounding its suitability as a training program (Lackey, Salcedo, Matthews, & Maxwell, 2014; Maxwell, 2015). Because it is unwieldy, not user-friendly, and lacks a common user-interface, participants had to spend time not only learning how to use the program, but also change any preexisting schemas to suit the requirements of VBS3. Individuals, firstly, will have to assign more cognitive resources to handle the extraneous load, as placing the soldiers in an unfamiliar system means that they will be unable to resort to using schemas to integrate the material. This then results in less cognitive resources available targeted towards developing schemas in repeated exposure to VBS3; the positive cycle of schema construction alluded to by Paas, Renkl, and Sweller (2003 p. 2) is thus obstructed by the heavy extraneous load caused by the program.

Additionally, because participants are required to integrate novel information to use the system, this hinders the ability to automate their actions, further impeding schema construction. As such, participants will experience an increased split-attention effect because they will have to simultaneously integrate the information presented simply to use the system, as well as actually pay attention to the serious game in order to be trained on the material. This is an undesired outcome of a training program, as the problematic allocation of cognitive resources has been shown to negatively affect training effectiveness due to a lack of consideration of cognitive architecture during training design. The effect of an inefficient cognitive burden may have affected a trainees' ability to learn and will be reflected in the training evaluations, demonstrating an undesired return on investment from the VBS3 program.

Figure 1 exhibits the typical set up station for a VBS3 training program with the included keyboard overlay acting as a legend. As mentioned, trainees had to learn and familiarize themselves with the complex setup both before and during the program. Therefore, an alternative serious game platform called the Military OpenSimulator Enterprise Strategy (MOSES) was developed and implemented with one of the purposes of simplifying the training program to optimize training effectiveness by mitigating the extra burden on extraneous cognitive load.

Figure 1. A typical Virtual BattleSpace 3 station.

Instead of having specific keys on the keyboard denoting certain actions, MOSES ran on simple point-and-click controls that did not require a reestablishment of schemas. Only minimally required graphics were included and all other unnecessary interfaces were omitted to mitigate unnecessary cognitive load (see Figure 2). Researchers observing the training program reported that VBS3 took twice as long on average for participants to pick up the controls as compared to MOSES.

Figure 2. A typical MOSES station.

Proposed Study

The purpose for this study was to investigate the relationships between the cognitive load, the type of system used during a training program (VBS3 vs. MOSES), and the performance measures from a training evaluation. Based on the concepts discussed and the difference in controls and user-interfaces between VBS3 and MOSES, it was hypothesized that the groups trained using the MOSES platform will perform better than the groups trained using the VBS3 platform on all performance measures. Furthermore, this group difference was hypothesized to be explained by the heavier cognitive load caused by VBS3, where cognitive load will act as a mediator between modality and performance measures.

Thus, the following hypotheses were examined:

$H_1 =$ Performance measures for groups trained in the MOSES serious game will be higher than groups trained in the VBS3 serious game.

$H_2 =$ Cognitive load will mediate the relationship between the condition of training program and the performance measures.

The concepts discussed also suggested that there may be a moderating effect by condition of training program on the relationship between cognitive workload and the resulting performance on the training evaluation activity. The instructional design of the VBS3 training program may cause a higher cognitive workload burden than the MOSES training program, and therefore because there will be a heavier extraneous cognitive load imposed through VBS3, less cognitive resources remain. As a result, the aforementioned cycle of schema construction and automation by Paas, Renkl, and Sweller (2003) is not supported by the design of the game, which will lead to consistently lower performance. Conversely, as cognitive workload increases through the use of MOSES, the extraneous load is still kept at a minimum, which will facilitate schema construction and automation, and augment continuous use of the game. Therefore, even though a heavier cognitive load will affect performance, the effect will not be as salient.

$H_3 =$ The condition of training program will act as a moderator on its indirect effect on performance measures through cognitive workload.

Figure 3 represents the conceptual framework for the process analysis model for the hypotheses that were examined in this study.

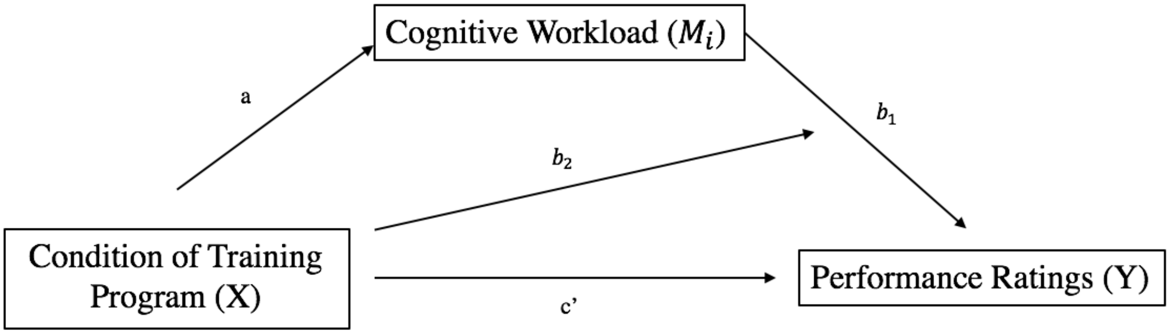


Figure 3. Conceptual conditional process analysis model.

CHAPTER THREE: METHODOLOGY

Overview

This study utilized data collected from a field study conducted at Camp Blanding Regional Training Institute with the 211th Florida Army National Guard between November 2015 and January 2017 under the Cooperative Agreement (CA) #W911NF-14-0012, between the University of Central Florida and the United States Army Research Laboratory. This CA was derived from a call for investigation into the affordances of serious games for training purposes, and fulfilled a need for Army research into comparing training effectiveness evaluations of serious game training programs and the more traditional "live" training programs for ensuring successful transfer of training (GAO, 2016). Approval from the Army Research Laboratory's Institutional Review Board (IRB) and the university's IRB were obtained prior to beginning the study (see Appendix E; see Appendix F).

Because military training is often demanding and fast-paced; it is uncommon for researchers to be able to acquire data from a study implemented in the field, and the researchers of this particular study were fortunate to have a cooperative unit in participation. Authorization from the 211th Regional Training Institute command staff and course managers were obtained, as well as participation agreement from course instructors and soldiers.

A monthly Basic Leadership Course cycle lasted 20 days, with days 1-17 in the classroom in the form of a traditional lecture, day 18 in the respective serious game training programs, which had been built to replicate the on-site situational training exercise (STX) lanes, and days 19-20 being assessed at the physical STX lanes. The STX lanes were developed with the purpose of training soldiers to exercise their decision-making skills, and focused largely on cognitive rather than physical abilities. This was optimal for our investigation into the role of cognitive load in different serious game platforms. As shown in Figure 3, tasks such as reacting to an improvised explosive device (IED) or an ambush required soldiers to decide how to respond immediately. Participants completed an informed consent form prior to beginning the Basic Leadership Course (see Appendix A).

During the serious game training session, soldiers were randomly assigned to either the VBS3 or MOSES conditions, and participated in several rounds of orchestrated, whole group, computer-based simulation scenarios that included face-to-face role play of leadership activities. Following completion of the computer-based simulation training, participants completed the NASA-TLX form. The live evaluation of training was conducted at the base's designated STX lanes, where they were assessed on four overall performance measure tasks. Throughout the STX lanes, squad group leaders rated each squad using the performance rating form. Participants of this study were divided into a total of 28 groups across the two conditions and were comprised of 250 soldiers. Missing data was excluded from analyses through listwise deletion, as the conditional process analysis method required complete data. However, data was missing completely at random and not a result of the condition. For example, missing data could be due to the attrition of one participant, resulting in the exclusion of the whole team from the data collection process.

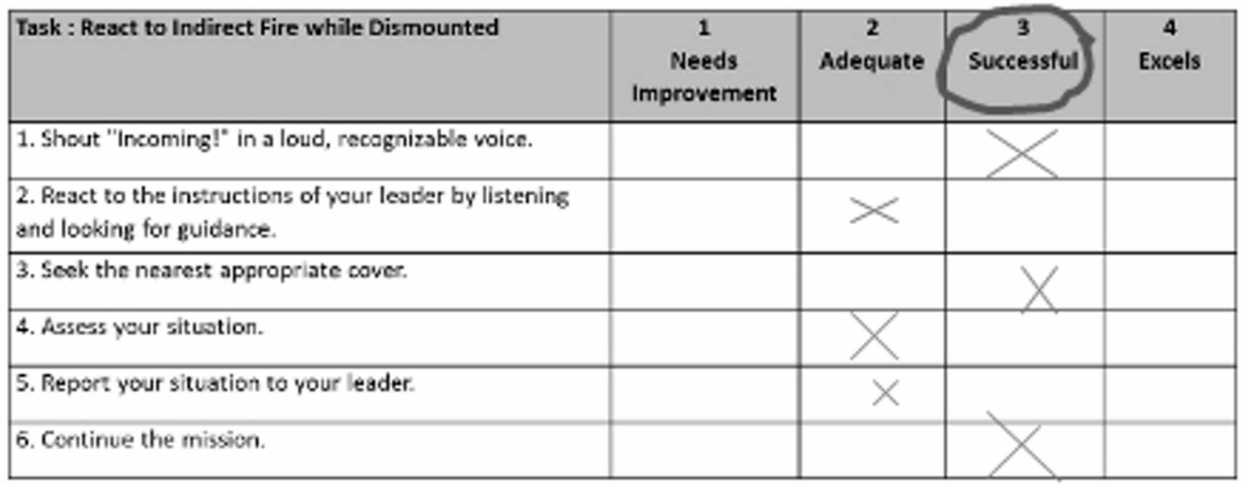
Figure 4. Soldiers at the STX lanes.

Measures

Performance Ratings

Participant performance ratings were based on the STX lane group ratings assessed by two course instructors, and consisted of four overall tasks (see Appendix B). Each task cluster included 4-6 tasks with a rating scale from 1-4, ranging from “Needs Improvement” to “Excels”. All items were presented in the form of a paper-and-pencil rating scale that were using during the STX lanes on days 19-20, and the rubric was adapted and described by Maxwell, Stevens, and Maraj (2016) where a binary score was replaced with proficiency levels.

Figure 4 portrays an example evaluation sheet of a task cluster; as demonstrated, scoring and analysis of data was comprised of two numbers for each variable: the overall score for React to Indirect Fire while Dismounted was 3 (in the grey highlighted cells) for the participating squad and the mean score of the participating squad was 2.5 (a mean of the six tasks included under this specific task cluster). Mean scores were included in the analyses because the overall instructor rating scores were restricted to the scale. In the case of the example evaluation sheet, each subtask under this overall task varied consistently between 2 and 3; yet the overall score allocated for this task was a 3. Therefore, tabulating a mean score of each task for each squad allowed for greater specificity of, and yielded more information about each squad's performance.



Task : React to Indirect Fire while Dismounted	1 Needs Improvement	2 Adequate	[✓ 3 Successful]	4 Excels
1. Shout "Incoming!" in a loud, recognizable voice.			X	
2. React to the instructions of your leader by listening and looking for guidance.		X		
3. Seek the nearest appropriate cover.			X	
4. Assess your situation.		X		
5. Report your situation to your leader.		X		
6. Continue the mission.			X	

Figure 5. Example data evaluation sheet.

Cognitive Load

Cognitive load is measured by the NSA-TLX (Task Load Index), which has been validated and well-established by Hart and Staveland (1988) as an appropriate measure of workload. The global measure includes six factors: mental demand, physical demand, temporal demand, performance, effort, and frustration (see Appendix C). Because this study was concerned with the cognitive aspect of load, the variable of interest for analysis was mental demand. Mental demand in particular was selected due to the design of the subscales of the measure allowing for an understanding of how overall workload is distributed among the six components. In the NASA-TLX measure, mental demand is worded as so: “How much mental and perceptual activity was required (e.g. thinking, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?”, which taps specifically and exclusively into the mental requirements of a task as opposed to the wordings of the other five subscales (Hart & Staveland, 1988). Additionally, past literature has concentrated on analyzing subscale ratings when examining various aspects of mental workload instead of a global workload scores, specifically mental demand (Hart, 2006; Vitense, Jacko, & Emery, 2003), and studies have confirmed a significant contribution of mental demand in perceived mental workload (DiDomenico & Nussbaum, 2008; Finomore, et al., 2006; Zhang, Ayres, & Chan, 2011). Therefore, mental demand was selected in an effort to limit the amount of analyses required and based on past literature.

Demographics

Participants completed a demographic questionnaire, which included questions such as age, sex, years in the military, ranking, etc., which was administered in a paper-and-pencil format (see Appendix D).

CHAPTER FOUR: FINDINGS

The data analyzed in this study were cleaned to ensure that there were no mistakes in data entry as well as to ensure that any missingness was completely random. Missing data was excluded from analyses through listwise deletion due to the requirements of conditional process analysis. Preliminary analyses were conducted on the data, and the conditional process model was analyzed through PROCESS (Hayes, 2017).

Demographics

Descriptive Statistics

Descriptive statistics of the data set were collected on the demographics. The mean, standard deviation, minimum, and maximum were calculated for participant's age. Frequency statistics was collected on sex. The mean and standard deviation for each sub-task score as well as overall task and overall task scores were calculated for each group.

The sample consisted of a total of 250 soldiers. The mean age was 27.26, with almost 90% between the ages of 19 and 48. There were a total of 199 (79.6%) male participants and 51 female participants (20.4%). This disproportionate distribution was expected and is representative of the population, as there is currently a larger percentage of males in the military than females. The mean amount of years in the military was 5.64 years, ranging from one to 16 years, where 227 (91.9%) participants had under 10 years in the military. The sample contained a high educational level, with around 80% of the respondents reporting having some college education or higher. More than half of the participants reported having some college education

(64.8%), 18.2% reported having a bachelor's degree, and 1.6% reported having a master's degree. Table 1 includes the demographic profiles of the participants.

Table 1: Demographic profile of participants

Variables	Category	Frequency	Percent (%)
Gender	Male	199	79.6
	Female	51	20.4
Age	19	1	0.4
	20	11	4.4
	21	11	4.4
	22	20	8.0
	23	27	10.8
	24	16	6.4
	25	25	10.0
	26	24	9.6
	27	15	6.0
	28	15	6.0
	29	13	5.2
	30	13	5.2
	31	9	3.6
	32	10	4.0
	33	6	2.4
	34	6	2.4
	35	5	2.0
	36	1	0.4
	37	4	1.6
	38	3	1.2
	39	4	1.6
	41	1	0.4
	42	2	0.8
	43	3	1,2
	45	1	0.4
	48	1	0.4
Years in Military	1	2	0.8
	2	12	4.8
	2.5	4	1.6
	3	42	16.8
	3.5	1	0.4
	4	38	15.2

Variables	Category	Frequency	Percent (%)
	4.5	2	0.8
	5	33	13.2
	5.5	3	1.2
	6	26	10.4
	6.5	2	0.8
	7	23	9.2
	8	30	12
	9	8	3.2
	9.5	1	0.4
	10	7	2.8
	11	4	1.6
	12	4	1.6
	13	1	0.4
	14	1	0.4
	15	1	0.4
	16	2	0.8
Education	GED	8	3.2
	High School	28	11.2
	Some College	160	64.0
	Bachelor’s	45	18.0
	Master’s	4	1.6
	Other	2	0.8

Preliminary Analysis

To provide a preliminary analysis for the distribution of the performance measures, histograms were generated for the cognitive load measures collected by the NASA-TLX survey, as well as performance measures for each sub-task score, overall task score, and average task score for all tasks. Descriptive statistics of the performance measures as well as mental demand were collected. The final sample size of group performance measures for overall ratings after listwise deletion of missing data was *n*=14. This was due to several squad group leaders unaware of the requirement of the “overall” rating score, which led to them only reporting the scores on each individual task under the task cluster. For example, in the example evaluation sheet depicted in Figure 4, the squad group leader would not have reported a total overall score of "3", and instead would have only scored a group on each individual task, such as "Shout 'Incoming!' in a loud, recognizable voice." Future research should ensure that squad group leaders are aware of the overall score rating as well.

The means of the overall and average scores on the four task clusters were between 2.44 (from React to Far Ambush Average) and 3.29 (for React to Indirect Fire Overall and React to Improvised Explosive Device Overall). Table 2 reports the means and standard deviations for each task cluster as well as the individual tasks for overall and average scores. The mental demand self-report scores ranged from 13.33 to 62.86 (*M* = 42.85, *SD* = 13.73).

Table 2: Descriptive statistics for overall and average task ratings

Task	*n*	Mean	Standard Deviation	Skewness	Kurtosis
React to Indirect Fire Overall	14	3.29	0.73	-0.52	-0.73
React to Indirect Fire Average	26	3.13	0.50	0.13	-0.98
Shout in Recognizable Voice	28	3.29	0.90	-1.29	1.20
React to Leader's Instruction through Guidance	28	3.13	0.72	-0.15	-0.99
Seek Nearest Appropriate Cover	27	2.48	1.05	0.16	-1.12
Assess Situation	28	3.07	0.66	-0.08	-0.54
Report Situation to Leader	28	3.11	0.69	-0.14	-0.72
Continue Mission	27	3.44	0.51	0.24	-2.11
React to IED Overall	14	3.29	0.61	-0.19	-0.26
React to IED Average	17	3.00	0.59	0.58	-0.71
Performed 0-2-25 Checks	22	2.55	1.10	-0.13	-1.25
Establish 360-degree Security	27	3.04	0.65	-0.03	-0.40
Tactical Combat Casualty Care Measures	22	2.64	0.85	-0.21	-0.29
Evaluate Casualties	23	2.91	0.60	0.01	0.16
Submit Situation Report to Higher HQ's	27	3.33	0.48	0.75	-1.56
Continue Mission IAW Higher HQ Guidance	26	3.42	0.50	0.33	-2.06
React to Near Ambush Overall	15	2.53	0.99	0.15	-0.84

Task	*n*	Mean	Standard Deviation	Skewness	Kurtosis
React to Near Ambush Average	28	2.65	0.70	0.14	-0.86
Return Fire Immediately	28	2.71	0.98	-0.14	-0.95
Kill Zone – Assault through Ambush	28	2.39	0.96	0.19	-0.77
Not in Kill Zone – ID Enemy, Suppressive Fire	28	2.25	0.97	-0.02	-1.15
Assault and Destroy Enemy Position	28	2.57	0.96	-0.08	-0.83
Unit Leader Reports Contact to Higher HQ	28	3.32	0.61	-0.29	-0.53
React to Far Ambush Overall	15	2.60	0.63	0.55	-0.39
React to Far Ambush Average	16	2.44	0.56	-0.05	-0.59
Receiving Fire – Return Fire, Seek Cover, Suppress Enemy	27	2.44	0.89	0.36	-0.49
No Fire – Concealed Route to Enemy Flank	27	2.56	0.75	-0.20	-0.05
FO Calls for and Adjusts Indirect Fires	16	2.38	0.89	-0.23	-0.65
Kill Zone – Shift Suppressive Fires	26	2.58	0.86	-0.26	-0.36

The performance measures consisted of four tasks with overall and mean scores collected for each task. To analyze the conditional process relationships between the variables of interest, eight different analyses would need to be run for each task cluster for both overall and mean scores. Therefore, a correlation analysis was conducted between the eight different performance outcome measures to limit the amount of analyses, where significant correlations will indicate similarity in variable relationships, signifying redundancy in analyses. Selection of final tasks depended upon the greatest number of significant correlations with other tasks. React to Near Ambush Overall correlated significantly with four tasks, and following tasks were selected based on significant correlations with other tasks that did not correlate with React to Near Ambush

Overall, as well as large standard deviation values (see Table 2). This was to ensure greatest variation in the performance measure ratings, which should allow for greater distinction between rating scale points. React to Improvised Explosive Device Overall and React to Far Ambush Average fulfilled this criteria and were selected for final analyses. Results of correlation analysis are included in Table 3.

Table 3: Correlation analysis of performance measures

Variables	1	2	3	4	5	6	7	8
1. React to Indirect Fire Overall	–							
2. React to IED Overall	-.03	–						
3. React to Near Ambush Overall	.62*	.19	–					
4. React to Far Ambush Overall	.41	.28	.02	–				
5. React to Indirect Fire Average	.80**	.19	.66*	.29	–			
6. React to IED Average	.36	.89**	.26	.53	.21	–		
7. React to Near Ambush Average	.64	.29	.95**	.02	.61**	.31	–	
8. React to Far Ambush Average	.44	.57	.85**	.76*	.49	.76**	.69**	–

*p < .05. **p < .01.

An investigation of possible group differences between the two conditions (VBS3 vs. MOSES) was conducted. For assessing group difference between VBS3 and MOSES, the tasks were tested for normality using Shapiro-Wilk test, P-P plots, and skewness and kurtosis were examined. Homogeneity of variance was checked with Levene's test. For React to Indirect Fire Overall, Levene's test indicated homogeneity of variance (*F*(1,12)= .86, *p*= .37), and Shapiro-Wilk test indicated non-normality (*p* = .001). Therefore, a Mann-Whitney U test was used to evaluate possible group differences. Analysis found that groups trained using MOSES (*Mdn* =

3.50) did not score significantly higher than groups trained using VBS3 (*Mdn* = 3.00, *U* = 14.00, *p* = .23, *r* = -0.39). For React to Near Ambush Overall, Levene's test did not find heterogeneity of variance (*F*(1,13)=.16, *p* = .70) and Shapiro-Wilk test was nonsignificant. Therefore, an independent samples t-test was conducted to compare group differences and found no significant difference for the MOSES (*M* = 2.75, *SD* = 1.04) and VBS3 (*M* = 2.29, *SD* = 0.95) conditions; t(13) = -0.90, *p* = .39. For evaluating the performance measure of React to Far Ambush Average, Levene's test was nonsignificant (*F*(1,14) = .67, *p* = .42), and Shapiro-Wilk test was nonsignificant; therefore, an independent samples t-test was conducted. Analysis found no significant differences between the group trained using MOSES (*M* = 2.46, *SD* = 0.61) and the groups trained using VBS3 (*M* = 2.38, *SD* = 0.43), t(14) = -0.25, *p* = .81.

Further investigation in possible group differences in mental demand between VBS3 and MOSES was conducted. Levene's test confirmed homogeneity of variance (*F*(1,20)=.003, *p* = .96), and Shapiro-Wilk test was nonsignificant. An independent samples t-test indicated no significant group differences between the MOSES (*M* = 42.91, *SD* = 13.16) and VBS3 (*M* = 42.70, *SD* = 16.49) conditions; t(20) = -0.03, *p* = .98.

While no significant differences were found between the MOSES and VBS3 conditions on performance measures and cognitive workload, conditional process analysis was still conducted on the model. Literature on mediation analysis has begun to depart from the Baron and Kenny method, emphasizing that the "effect to be mediated" (Judd & Kenny, 1981; Preacher & Hayes, 2004), or requirements of significant relationships between the variables, is a considerable limitation (Pardo & Román, 2013; Zhao, Lynch, & Chen, 2010). The absence of an effect to be mediated could be due to a variety of reasons (such as covariates, time intervals, or designs that are lacking statistical power to detect an effect), but that does not necessarily indicate a nonsignificant mediation relationship.

Conditional Process Analysis

According to literature, the bootstrap method, a non-parametric resampling test by Preacher and Hayes (2004), is robust to violations of assumptions, and is ideal for studies with smaller sample sizes (Erceg-Hurn & Mirosevich, 2008; Vallejo Seco, Ato García, Fernández García, & Livacic Rojas, 2013; Hadi, Abdullah, & Sentosa, 2016; Imai, Keele, & Yamamoto, 2010). Therefore, conditional process analysis was used to analyze the mediation of variables of interest, with the condition of training program (VBS3 or MOSES) as the exogenous variable (X), cognitive workload (mental demand) as the mediating variable (M), and performance measures as the consequent variable (Y) (see Figure 3). Furthermore, the condition of training program was analyzed as a moderator variable between cognitive workload and performance measures, which caused a conditional indirect effect (Hayes, 2017). Judd and Kenny (1981) described this conditional process model, in which the causal variable (X) is able to moderate its indirect effect on the outcome variable (Y) through M because it acts as a moderating variable on the relationship between M and Y. This model has also been referred to as the original conditional process model.

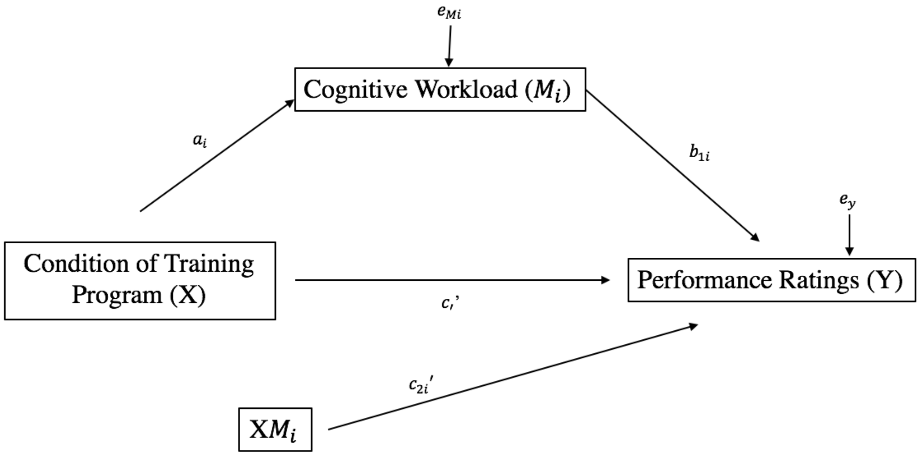


Figure 6. Statistical conditional process analysis model.

From the statistical form of this model shown in Figure 6, we were able to extract the two linear equations of the model:

$$\mathrm{M_i = i_1 + a_iX + e_{M_i}} \qquad (1)$$

$$\mathrm{Y = \; i_2 + c_1{}'X + b_{1i}M_i + c_{2i}{}'XM_i + e_y} \qquad (2)$$

The equation for the consequent *Y* was rewritten as:

$$\mathrm{Y = \; i_2 + c_1{}'X + (b_{1i} + c_{2i}{}'X)M_i + e_y} \qquad (3)$$

Where the conditional indirect effect on *Y* through *M* was:

$\mathrm{a\theta_{M\rightarrow Y} = a(b_{1i} + c_{2i}{}'X)}$ (4)

PROCESS was used to analyze the data and conduct conditional process analysis with 5000 bootstrap samples on the three selected performance measures. Performance measures were based on overall group performance and cognitive load measures were collected on an individual basis; therefore, the mean cognitive load measure for each squad was aggregated based upon the rationale from Bowers, Braun, and Morgan (1997): due to the low interdependence of the various roles within each group during an evaluation, because a successful performance does not rely solely upon one individual but on an equal contribution from each group member, team workload measures can be gathered by calculating a mean of the individual scores.

For the performance measure of React to Improvised Explosive Device Overall, the conditional process analysis of mediation indicated no significance (Table 4), and with a nonsignificant conditional direct effect of moderation (*p* = .22). For the performance measure of React to Near Ambush Overall, conditional process analyies also indicated nonsignificance in mediation between the variables in the model (Table 5), and nonsignificance in conditional direct effect of moderation (*p* = .35). Finally, for the performance measure of React to Far Ambush Average, analysis resulted in a nonsignificant mediation of the variables in the model (Table 6), and a nonsignificant moderating relationship (*p* = .50).

Table 4: Results for the model (PROCESS Model 74, Hayes, 2017) investigating conditional process relationship between condition of training program, mental demand, and performance measure for React to Improvised Explosive Device Overall.

Outcome Measure	Predictors	Coefficient (SE)	*t*	LLCI, ULCI
Mental Demand	Constant	0.00 (4.78)	0.00	-10.66, 10.66
	Condition of Training Program	-4.56 (9.56)	-0.48	-25.88, 16.75
	$F(1,10) = 0.23, p = .64, R^2 = 0.02$			
React to IED Overall	Constant	3.25 (0.19)	17.54	2.82, 3.68
	Mental Demand	0.01 (0.01)	0.81	-0.02, 0.04
	Condition of Training Program	0.55 (0.37)	1.47	-0.31, 1.40
	Mental Demand X Condition of Training Program	-0.001 (0.02)	-0.04	-0.06, 0.06
	$F(3,8) = 0.85, p = .50, R^2 = 0.24$			

Table 5: Results for the model (PROCESS Model 74, Hayes, 2017) investigating conditional process relationship between condition of training program, mental demand, and performance measure for React to Near Ambush Overall.

Outcome Measure	Predictors	Coefficient (SE)	*t*	LLCI, ULCI
Mental Demand	Constant	0.00 (4.78)	0.00	-10.66, 10.66
	Condition of Training Program	-4.56 (9.56)	-0.48	-25.88, 16.75
	$F(1,10) = 0.23$, $p = .64$, $R^2 = 0.02$			
React to Near Ambush Overall	Constant	2.81 (0.24)	11.89	2.27, 3.36
	Mental Demand	-0.01 (0.02)	-0.32	-0.04, 0.03
	Condition of Training Program	0.48 (0.47)	1.01	-0.61, 1.57
	Mental Demand X Condition of Training Program	0.05 (0.03)	1.72	-0.02, 0.13
	$F(3,8) = 1.53$, $p = .28$, $R^2 = 0.36$			

Table 6: Results for the model (PROCESS Model 74, Hayes, 2017) investigating conditional process relationship between condition of training program, mental demand, and performance measure for React to Far Ambush Average.

Outcome Measure	Predictors	Coefficient (SE)	*t*	LLCI, ULCI
Mental Demand	Constant	39.69 (9.33)	4.25	19.36, 60.03
	Condition of Training Program	5.96 (10.53)	0.57	-16.98, 28.91
	$F(1,12) = 0.32$, $p = .58$, $R^2 = 0.03$			
React to Far Ambush Average	Constant	2.40 (0.74)	3.25	0.76, 4.04
	Mental Demand	0.005 (0.02)	0.28	-0.03, 0.04
	Condition of Training Program	-0.68 (0.97)	-0.70	-2.84, 1.48
	Mental Demand X Condition of Training Program	0.01 (0.02)	0.49	-0.07, 0.06
	$F(3,10) = 0.53$, $p = .67$, $R^2 = 0.13$			

CHAPTER FIVE: DISCUSSION AND CONCLUSION

Summary

The purpose for this study was to investigate the role of cognitive load in the relationship between the condition of training program and outcome on performance measures. Data that were collected from a field study between November 2015 and January 2017 were checked for accuracy in SPSS, and screened for missing or incorrect information, skewness, kurtosis, and homogeneity. Descriptive statistics were generated for an overview of participants' profiles, and quantitative analyses were conducted to test the structural model. Correlation analysis of the eight performance measures were conducted to identify tasks that were highly correlated with other tasks in order to limit the analyses. Final tasks were chosen by greatest number of significant correlations with other tasks as well as larger values of standard deviations, which resulted in the selection of the tasks: React to Improvised Explosive Device Overall Rating, React to Near Ambush Overall Rating, and React to Far Ambush Average Rating. A Mann-Whitney U test was used to investigate group differences between the condition of training program on performance measures for preliminary analysis. Data analysis testing the proposed model was conducted using PROCESS analysis (Hayes, 2017) in SPSS, and resulted in nonsignificant findings between the variables.

Discussion

The research question that drove this investigation was an inquiry into possible negative effects of technology in training programs. Due to the pervasive nature of the "transfer problem", ensuring training effectiveness should always remain in the forefront of the training industry. However, this study did not find results that supported the hypotheses that cognitive load would explain the difference in performance measures due to training condition (VBS3 or MOSES), and that the training condition would moderate the relationship between cognitive load and performance measures. There are several avenues to explore that may offer an explanation for the results of this study.

First, the notion that there would be a difference in performance measures between the two conditions was a foundational premise of this study; the structure of the model was predicated on the belief that VBS3 imposed a higher cognitive load than MOSES. However, conditional process analyses of the three selected tasks resulted in nonsignificant differences; therefore, the first exploration focuses on the source of the foundational premise of this study, and why the study was initially proposed. The manipulation of instructional design to reduce extraneous cognitive load is not a novel idea; Sweller's introduction of cognitive load theory in the late 1980s led to an era of investigative efforts to minimize the proposed cognitive load by aligning instructional design of materials to cognitive architecture (Paas, Renkl, & Sweller, 2004). With the advancement of technology and its steady incorporation into training and learning contexts, this investigative era must be revisited, and experts have recommended to "simplify and organize the delivery of information to allow for encoding in visual and verbal channels, and to improve opportunities for mental model building" (Nelson & Erlandson, 2008, p. 637) when studying cognitive load caused by synthetic learning environments.

The coherence principle, posited by Nelson and Erlandson (2008), refers to the omission of extraneous material that is irrelevant to the information for learning or training purposes, in order to present a more coherent message that may facilitate deeper learning. However, this endeavor may not always be achievable in the realm of immersive and visually complex environments, as material that could be omitted may be inherent to the immersive experience in the environment. It may be difficult to explicitly measure the extent to which a graphical item in the immersive environment adds to the user's experience, and if its omission will encourage or prohibit deeper learning. Therefore, while it is impossible to collect explicit measures of whether synthetic learning environments align with the coherence principle, it is still possible, and important, to investigate the cognitive load imposed upon a user by the environment.

Furthermore, while there have been no prior studies examining the difference in cognitive load between the VBS3 and MOSES platforms, several studies exploring cognitive load with NASA-TLX have resulted in finding high workload from using VBS3. For example, an experiment by Lackey, Salcedo, Matthews, and Maxwell (2014) found a higher significant difference in the post-training NASA-TLX mental demand scores for the groups trained in VBS3 than groups trained in the live condition. Reinerman-Jones, Goodwin, Goldiez, and Wismer (2017) found that for the condition of VBS3, "workload was high in terms of mental demand, effort, and frustration" (p. 7). Additionally, a study investigating the evaluation and testing of virtual environments using the mental demand measure from NASA-TLX stressed that "VBS3 is often criticized for being difficult to deploy and use, however, actual usage data is difficult to acquire" (Maxwell & Zheng, 2017, p. 1). While obtaining data that implicates VBS3 as the source of unnecessary extraneous cognitive load remains challenging for researchers, literature has shown that using VBS3 as a training program results in high mental demand through the measure of cognitive load. Perhaps the assumption that MOSES would impose a significantly lower cognitive load than VBS3 was incorrect and all efforts to incorporate synthetic learning environments should be thoroughly investigated before implemented. Further research on this topic could impart valuable insight into maximizing the implementation of synthetic learning environments in training contexts.

An interesting possible explanation for the lack of group differences between the performance measures, as was assumed by the model of this study, could have been due to a caveat of the redundancy effect, where information including redundant material may result in less learning than if it were omitted (Mayer, 2001; Sweller, Ayres, & Kalyuga, 2011). Research has found that having redundant information may improve learning (Kalyuga, Chandler, & Sweller, 1999; Kalyuga, Chandler, & Sweller, 2000), and a study by Mayer and Johnson (2008) reported that if the redundant information is short and near the information it refers to, it may have no negative effects or even be advantageous for the individual’s learning. It was not possible to explore the boundaries of the redundancy effect and whether it was present in this study, but perhaps the redundant information presented by VBS3 that was hypothesized to cause extraneous cognitive load was short and near enough to the learning material for it to have not impacted participants’ learning.

The second avenue explores the validity behind cognitive load theory and possible explanations that may illuminate the nonsignificant findings. Although cognitive load theory is a well-established and extensively researched concept, because it was another key aspect of the model, it is pertinent to consider any misconceptions and misinterpretations that may have affected the accuracy of the predictions. Literature has revealed several complications in the application of cognitive load theory, and is worthwhile to investigate in relation to this study.

First, it is important to consider the theoretical soundness of cognitive load theory. Because cognitive load theory is based off several assumptions which cannot be confirmed empirically and therefore is not falsifiable, it must be interpreted with care (Gerjets, Scheiter, & Cierniak, 2009). The foundation which cognitive load theory is based upon is somewhat circular and exemplifies a structuralist approach of theory validation, in which its concepts cannot be presented and tested without the involvement of prior assumptions. Therefore, each new investigation of the application of cognitive load theory is a holistic test of not just assumptions that relate specifically to the study in question, but also of the existing assumptions of the theory behind cognitive load. Consequently, if a prediction of a study is not supported, the structuralist approach to deciphering the findings is a recommendation of not a rejection of the foundational assumptions of the theory, but instead, a modification of the specific assumptions to the particular measurement or instructional format.

According to this line of reasoning by Gerjets and colleagues (2009), the findings of this study should be interpreted as a misalignment of the assumptions to its measurement or instructional format (e.g. the assumption that the difference in the system setup between VBS3 and MOSES would cause a differing extraneous cognitive load that is salient enough to be pronounced in the selected performance measures). Consequently, future research into this topic should surround modifications of the design, measurement, and instructional format.

An alternative question that requires exploration is the inclusion of cognitive load theory in the model used in this study. The inclusion was due to the notion that if an individual does not possess the foundational knowledge to use the system in a learning environment, instructional guidance can serve as a substitute for the lack of schema associated with the system (Schnotz & Kürschner, 2007). However, this notion is based off of an absolute certainty in the understanding of cognitive load theory and its division of intrinsic, extraneous, and germane cognitive load, and many common instructional techniques may actually hinder a user's learning due to a dogmatic acceptance of the theory. Cognitive load theory may have been overgeneralized in its application to instructional design, and warrants further analysis; the idea behind simply decreasing extraneous cognitive load to increase germane cognitive load for maximal learning is perhaps more complex than current literature has demonstrated.

For instance, a possible complication in the theory that has been proffered is the possible indivisible nature of extraneous and germane cognitive load, resulting in a lack of returns from the instructional design of a system (De Jong, 2010). For this study, the intention behind the incorporation of MOSES was the hypothesized decrease in extraneous cognitive load afforded by the instructional design by removing unnecessary details and graphics, and by selecting a more user-friendly interface of the system. However, this reduction in extraneous cognitive load may have also resulted in the removal of affordances for germane processing; the presence of extraneous cognitive load may have stimulated germane cognitive load processes if the translation from one element to another (e.g. from the learning material to the legend) elicits deep knowledge processing. Therefore, perhaps for the participants in the VBS3 condition, the act of maintaining the information of their intended actions while translating from the learning material to the legend may have been beneficial by facilitating a deeper processing of actions in the decision-making process. By removing this required action for the MOSES condition, it is possible that the affordance for deep processing of their actions was also removed and therefore the effects of both conditions were neutralized, resulting in a lack of group difference.

Another theoretical issue with cognitive load theory that has been explored widely is the misalignment of learner expertise with task difficulty and instructional design. The simplistic application of reducing extraneous cognitive load to increase germane cognitive load is complicated by the differing levels of learner expertise across the population of all users and its influence on how effective or ineffective the material is for learning. If a learner has high expertise on the task at hand, his or her capacity will not be challenged, and if there is no challenge presented, learning may be impeded (Schnotz & Kürschner, 2007). Furthermore, a learner with high expertise does not require instructional facilitation. There is a possibility that the processing of the unnecessary instructional guidance imposes extraneous cognitive load, because the learner has to integrate the irrelevant information, which may require more time or effort without the extra benefit for learning. A study on learner expertise found that high-expertise learners had deeper comprehension of the text of the material, and did not require a high-coherence text (which was designed to decrease extraneous cognitive load) to facilitate their learning (McNamara, Kintsch, Songer, & Kintsch, 1996). This further strengthens the argument that presenting a task that has been simplified to a learner with higher expertise results in a decrease in challenging that learner's capacity, which may not facilitate, or even hinder his or her learning.

Therefore, although prior research on cognitive load has resulted in the determination that extraneous cognitive load should always be minimized, it seems as though there are instances in which it may be beneficial to align a learner's expertise with the task difficulty instead of always erring on the side of minimal extraneous load. Schnotz and Kürschner (2007) further support this argument by referring to Vygotsky's (1978) zone of proximal development, in stating that the instructional design should endeavor to stay within the zone, where the facilitation does not cause extraneous cognitive load, but is maximized and remains helpful for the learner's expertise level. It is possible that for MOSES, the instructional design that attempted to reduce extraneous cognitive load was, in actuality, instructionally negative facilitation, in which the facilitative nature of the instructional design reduces a task difficulty too much, to the degree that it is no longer aligned with the expertise level of the learner (Schnotz & Kürschner, 2007; Kalyuga, Chandler, & Sweller, 1998). Perhaps the instructionally negative facilitation in MOSES may have been manifested in the form of oversimplifying the game, and it may have no longer fit with a user's schema of a game. The point-and-click controls may have been too simple and offered no learning potential.

Kalyuga, Ayres, Chandler, and Sweller (2003) provided additional support for ensuring alignment of material to the level of expertise of the user by concluding from a study that the integration of multiple sources of information (as an instructional design to minimize the split-attention effect) lost its benefits as a learner's expertise increased, and slowly became disadvantageous. The authors argued that the integration of multiple sources of information is only beneficial for novices, and is redundant or even detrimental to experts. This expertise-reversal effect, where "a cognitive load that is germane for a novice may be extraneous for an expert… [where] information that is relevant to the process of schema construction for a beginning learner may hinder this process for a more advanced learner" (Paas, Renkl, & Sweller, 2004, p. 2) could explain the findings of this study. It is possible that the expertise levels of the participants were underestimated, and therefore the cognitive load that was intended as germane for the beginning learners had an expertise-reversal effect where it had no effect (or a minor, nonsignificant hindrance) on participants' learning processes because they did not require instructional help. The participants in the VBS3 condition did not receive this instructional facilitation, and therefore did not experience this expertise-reversal effect, and participants in the MOSES condition may have instead been neutrally or (nonsignificantly) negatively affected, resulting in the absence of group differences on the variables in question.

Alternatively, Kalyuga (2007) proposed that "instructional methods for enhancing levels of germane load may produce cognitive overload for less experienced learners, thus effectively converting germane load for experts into extraneous load for novice learners" (p. 515). It is also possible that the intended facilitative effect of the instructional design had a negative effect because the expertise levels of the participants were overestimated, and in this case imposed an extraneous cognitive load on MOSES users. It is difficult to identify the specific level of expertise of the participants in this study, and based on literature, possible that the efforts to reduce extraneous cognitive load has been the cause for the lack of effect, or negative effect on the participants, leading to a mitigation of any expected results.

Another idea driving this study was based on the theoretical additive nature of cognitive load; because of this assumption, decreasing extraneous cognitive load should lead to an increase in germane cognitive load. However, literature has revealed concerns with this proposed relationship, and it is important to consider the role of germane cognitive load in relation to intrinsic and extraneous cognitive load, as well as the limits of working memory. Kalyuga (2007) discussed the interchangeability of germane cognitive load, where it is only advantageous until it exceeds the working memory capacity, which at that point it is then regarded as extraneous cognitive load because it no longer facilitates learning. It is possible that the germane cognitive load imposed by MOSES for schema construction may have exceeded the working memory capacity and was converted to extraneous cognitive load, impeding learning. Thus, the intention behind promoting germane cognitive load by instructional design of MOSES may have hindered learning and negated any expected group difference.

A theoretical issue concerning the definitions behind mental load, mental effort, and performance was also posited as a plausible reason for the findings of this study. According to Kirschner (2002), “mental load is the portion of [cognitive load] that is imposed exclusively by the task and environmental demands. Mental effort refers to the cognitive capacity actually allocated to the task. The subject’s performance, finally, is a reflection of mental load, mental effort, and… causal factors” (p. 4). It would seem as though while mental load is a factor of the instructional design, mental effort is determined by the learner’s purposeful and conscious allocation of cognitive resources towards learning processes. Moreover, Jong (2010) asserted that intrinsic and extraneous cognitive load refer to cognitive activities that are required to be performed, and thus count as cognitive load. Germane cognitive load, on the other hand, refers to resources that have been left unused and therefore up to the learner's discretion and decision to manage, and can be categorized as mental effort.

The concept of germane cognitive load requiring conscious allocation has been examined by other experts in the field of cognitive load theory as well. Winne and Hadwin (1998) concluded that simply the reduction of cognitive load by task facilitation increasing working memory capacity does not automatically result in its allocation towards cognitive activities that augment learning. Learners do not generally engage in higher order cognitive processes that lead to greater comprehension of material, and therefore, that freed working memory is frequently left unused. A deeper investigation by Schnotz and Kürschner (2007) led to the identification of activities that classify cognitive processes as germane cognitive load: a purposeful application of learning strategies (that have not yet been automated), a purposeful identification of patterns in the material that will allow for schema acquisition, restructuring of mental representations of problems that allow for easier performance of a task, and metacognitive processes that actively monitor learning and cognition. Thus, it is clear that for cognitive load to be categorized as germane, there is a motivational aspect involved, and requires purposeful application. A learner has to consciously allocate the available mental resources towards strategic cognitive processing, an action that is influenced by factors such as an individual's learning processes, self-regulatory processes, and his or her motivational and affective aspects of learning.

For this study, if MOSES did decrease extraneous cognitive load as hypothesized, and allowed for greater resources to be allocated towards germane cognitive load, this does not necessarily mean that the participants consciously engaged in higher order cognitive processing such as schema acquisition and schema automation. Therefore, although MOSES may have had greater affordances, participants may not have been motivated enough to allocate unused resources towards germane cognitive processing, which may explain the lack of group differences in performance measures. Additionally, motivational aspects such as metacognitive processing and intrinsic motivation were not considered because the participant pool consisted of soldiers. While the study was voluntary, participants were required to follow instructions given by squad group leaders and did not have free-will within the synthetic learning environment to explore and develop characteristics that may have encouraged further motivation.

Schnotz and Kürschner (2007) also asserted that since learning is possible without working memory, and therefore without imposing germane load, it then follows that germane load is not a requirement for learning. Schema construction is not necessarily a conscious process that requires the involvement of working memory. This may explain the lack of moderating effects of the condition of training program between cognitive workload and performance measures, as well as absence of group differences in the variables of interest. Because working memory may not be associated with germane cognitive load as postulated by the authors, which was the basis of the theory behind the model used in this study, participants in the VBS3 condition may have been engaging in schema construction as well, without being hindered by the hypothesized difference in affordances for germane cognitive load.

In the further exploration of the allocation of germane cognitive load, Gerjets and Scheiter (2003) found no support for a hypothesized difference in results due to time pressure, where the authors predicted that the increase in time pressure would hinder germane cognitive load. This finding was rationalized by assuming that the learners made strategic decisions by choosing to allocate more resources towards germane cognitive load, and decreasing extraneous cognitive load by ignoring distracting information. This demonstrates that under the theoretical assumptions of the distribution of the different types of cognitive loads, the allocation of cognitive resources is up to the learner's decision and motivation. For this study, it is possible that the users of VBS3 were able to ignore the extraneous cognitive load caused by the system, such as unnecessary graphics and interface, and strategically allocate more towards germane cognitive load. This strategic allocation may then explain the lack of group differences in both mental demand and performance measures.

The most significant theoretical flaw about cognitive load theory that has been revealed throughout this investigation of current literature is the inability to measure specific types of cognitive load, therefore making it difficult to avoid the circulatory, post-hoc explanation of the division of types of cognitive load. This has been a common theme that has surfaced from all research and applications of cognitive load theory, and has been identified by various experts in this field. De Jong (2010) states that there is no clear distinction in the measurement between intrinsic, extraneous, and germane cognitive load; therefore, when one condition shows a difference in cognitive load, this may be a result of a lower extraneous load, lower germane load, or both. Schnotz and Kürschner (2007) further discuss this topic by giving an example: if a reduction in cognitive load has led to greater learning, then it can be concluded that it was due to a reduction in extraneous cognitive load; likewise, if there was an increase in cognitive load which led to an increase in learning, it can be concluded that it was due to germane load. These explanations cannot be claimed as explanatory, however, and although research has distinctly delineated the different types of cognitive load and shown support for their existence, the lack of explicit measurement obstructs the ability to make any authoritative assertions.

Gerjets and Scheiter (2003), as mentioned previously, found no difference in performance due to the hypothesized influence of germane cognitive load, and concluded that it was due to the strategic allocation of cognitive resources towards germane cognitive load. Participants were said to have acted accordingly to the theories of action control whereby enhanced task difficulty may have led to increased effort spent on the current task, which facilitated a decrease in vulnerability against distractional effects. For this study, it could be interpreted that participants in the VBS3 condition invested more effort into the current task due to enhanced task difficulty. This caused them to be less vulnerable to the extraneous effects of the system, such as having multiple legends (on the keyboard and a reference sheet below the system), or the additional graphics and user interface systems that were omitted from the MOSES condition. Comparatively, our initial hypothesis focused on the decrease in extraneous cognitive load as afforded by the MOSES system and its influence on outcome variables. Thus, the circulatory post-hoc nature of cognitive load is demonstrated.

This issue is clearly both prevalent and problematic as it leads to assumptions that cannot be explicitly proven, only deduced. To reiterate and bring this discussion full circle, this study found no group difference between the performance outcome measures or cognitive load. Therefore, if we are to apply the same circulatory post-hoc explanation to this study, we could assert that the lack of group differences in mental demand reported was due to participants in the MOSES condtion experiencing a heavier germane cognitive load than extraneous cognitive load, and participants in the VBS3 condition experiencing a heavier extraneous cognitive load than germane cognitive load. We could also assert that because mental effort is based off some motivation aspect, that there were differing levels of extraneous cognitive load for each condition, but equalizing levels of germane cognitive load as decided by the participating soldiers due to their decision of amount of resources to allocate. Additionally, we could infer from Gerjets and Scheiter's (2003) work that participants in VBS3 were able to strategically allocate more resources towards germane cognitive load and ignore the extraneous cognitive load caused by the system, resulting in a lack of mediation and moderation effects, and an absence of group difference on performance measures as hypothesized.

This review of the existing literature demonstrates a concerning aspect of cognitive load theory, and may provide clarifications for the findings of this study. The ambiguity behind the circulatory post-hoc explanations of the allocations of the types of cognitive load seem to underlie the other existing issues with the theory, such as the indivisible nature of extraneous and germane cognitive load, or the proposed influence of motivation on germane cognitive load, and brings into question the infallibility of the theory that this study was based upon. Furthermore, it is possible that while literature has identified VBS3 as the cause of high mental demand and high cognitive load, the effect may not be salient enough to have elicited a group difference.

It is also plausible that the nonsignificant findings of this study was due to methodological flaws in the design. There may have been a selection bias between the conditions simply due to the nature of random sampling and the possibility of groups that were not proportionately random. This may have led to a higher overall expertise level, for instance, of the groups in the VBS3 than the groups in the MOSES, which may have threatened the internal validity of the study. Expertise level may therefore have been a confounding variable and affected the influence of cognitive load by introducing a misalignment of expertise level to task difficulty for the VBS3 condition.

Furthermore, because this was conducted as a field study, it was difficult to gather a sample size large enough to ensure statistical power of the analyses. Participation depended on the availability of the soldiers and avoiding conflicts with scheduling. Researchers also needed to work together with instructors to prevent interfering with soldier training, as well as collaborate with the agendas of the instructors in accordance with the Basic Leadership Course. These complications were a challenge in the endeavor to increase sample size for this study. It is possible that a group difference did indeed exist, and there was a mediating and moderating relationship between the variables, but that the sample size was too small and the statistical power was too low to detect any differences. Fritz and MacKinnon (2007) assert that a sample size of 36 is the minimum required to detect large effect sizes in mediations, and a sample size of 558 is required to detect small effect sizes; comparatively, this study only consisted of 28 groups, which therefore this could have explained the results. It was also almost impossible to ensure reliability of treatment implementation. Because this was a field study and implementation was up to the squad group leaders' discretion, researchers were unable to intervene in the middle of a session. Therefore, if a squad group leader deviated from the study protocol or implemented the training program differently than planned, researchers were not able to interfere.

Due to the nature of field studies, this study was also susceptible to random irrelevancies in the setting, which may have introduced other confounding variables. Having complete control over the environment of the experiments was an impossible expectation due to the instructors having an agenda that required completion within a certain time period. Therefore, participants were not assigned the same squad group leaders across all groups, which may have increased or alleviated workload experienced. Additionally, the classroom setups were also not standardized across and within the conditions due to the requirement of having each group complete the training program within the assigned 20 days. Participants may have been exposed to differing environments and classroom setup positions, where one group may have been facing each other, and another group may have been facing the front of the classroom. As with all research experimentations, it is always ideal to ensure complete standardization across and within the groups to identify specific factors that influence the variables of interest; however, because of the lack of consistency of experimental setting, this may have introduced other factors that may have affected workload and performance measures.

Researchers also encountered difficulty ensuring reliability of the measures. Because the squad group leaders were not given any rater training for the performance assessments conducted at the STX lanes, there may be inaccuracies in the performance measures. For example, squad group leaders were not given frame-of-reference training; as a result, there may be large discrepancies between the ratings of two separate groups by different squad group leaders, and groups within one condition may have varying scores with large measurement errors. Raters were also not informed about the possible biases, and may have been vulnerable to the halo bias, for instance, by rating groups high on all the four tasks because they performed well on one task. As reported, the means across the performance ratings were between the mid “2’s” and low “3’s”, where squad group leaders tended to score groups close to the midpoint or higher, perhaps as a result of lack of rater training. Therefore, due to the unreliable nature of the data, it was not possible to detect regression as the rating scale on the performance measures did not allow for differentiation in performance. It could be due to squad group leaders viewing the score of a “1” as “bad” performance, and therefore avoiding that score purposefully across both VBS3 and MOSES conditions. Including a section discussing that it is acceptable to rate groups “1” on performance in a rater training session may have prevented this issue. It is also possible that the rating scale did not offer enough points on the scale to distinguish clearly between performance of the participants. This was, again, demonstrated by the low variance between the performance measures with the means fluctuating between the mid “2’s” and low “3’s”. Future research on this topic should also consist of the creation and validation of a measure with more discriminating points on the scale.

Furthermore, research has indicated that while the NASA-TLX scale has been a well-validated measure for cognitive load, it is still disadvantageous as a subjective rating measure. Individuals are often unable to accurately state the amount of cognitive load they are experiencing, and the measures are often a reflection of the instability of an individual’s framework of reference, which could fluctuate across the training program due to processes of adaptation, or emotional and motivational changes (De Jong, 2010; Schnotz & Kürschner, 2007). Thus, individuals themselves frequently are a source of unreliability of measurement as well due to the self-reporting nature of the scale. If the participants were unable to reliably report their cognitive load, this would have influenced the accuracy of the scores and ability of the scale to reflect any group differences between the VBS3 condition and MOSES condition.

Implications

This study has several implications that are important to both theoretical and practical applications. The results of this investigative journey highlighted the ambiguity of the application of cognitive load theory to research endeavors, as well as supported the importance of being cognizant of contextual factors that may influence technological implementation and use.

First, it is critical to note that although this study did not find significant results that supported the hypotheses, there were various limitations and as well as interpretations of the theory supporting the model that may have explained this outcome. However, this does not discount the significance of investigating factors in instructional design that may both maximize and hinder learning when incorporating technology in training programs. As technology becomes more prevalent in industries such as organizations, education, and the military, its introduction should always be driven and preceded by a focus on effective implementation and evaluation.

Theoretical Implications

A theoretical implication of this study is derived from the various caveats of cognitive load theory. While this theory has been grounded in decades of research and is well-established, this study supports the assertion that there are numerous factors that may impede successful research. As postulated by Gerjets, Scheiter, and Cierniak (2009), each exploration of a specific research idea and its relation to cognitive load is in actuality a holistic investigation of the theory and its assumptions as well, and not simply its specific application to that particular idea. It thus follows that should future research endeavors result in unsupported findings, researchers should thoroughly investigate the research design and instrumental application for possible influential factors as well as strategies for improvement.

Additionally, this study demonstrated a requirement for further examination of the theoretical relationship between the NASA-TLX scale and cognitive load. As literature reports, there is a difference between mental load and mental effort, where mental load is imposed upon the system and required by the task (i.e. intrinsic and extraneous cognitive load) and mental effort is decided by the learner (i.e. germane cognitive load). It is important to ensure that when investigating the different types of cognitive load, the researcher is explicitly clear in both constructing the research design as well as when conveying information to the participants the type of information he or she is hoping to gain through using the NASA-TLX. Based upon the implications of the differences between mental load and mental effort, the results could vary drastically if they are misunderstood. For instance, if the researcher is interested in studying germane cognitive load through measuring mental effort, but the participant interprets the “mental effort” item on the NASA-TLX to mean the load imposed by the system, the participant would be reporting either intrinsic or extraneous cognitive load. This mistake may threaten the strength at which the researcher can determine a relationship between the variables of interest. Therefore, this could be resolved with validated measures of intrinsic, extraneous, and germane cognitive load. While research has been making progress on this endeavor, it should remain a constant focus in the cognitive load field.

Practical Implications

The results of this study provide various practical implications. While this study did not investigate motivation, literature on this topic has affirmed its relationship to the allocation of cognitive resources and desired learning outcomes. It may benefit learners if the instructional design is not just focused on decreasing extraneous cognitive load, but also builds environments that are conducive towards increasing a user's motivation within the program. This begs the question of what techniques can be employed to accomplish this goal efficiently: for example, allowing time for learners to explore the affordances of the environments. In an examination of determinants of a user's perception and acceptance of a synthetic learning environment, Venkatesh (2000) found that computer playfulness, which relates to intrinsic motivation, involves not only a desire for fun, but also exploration and discovery within the environment, as well as challenge and curiosity.

Therefore, if learners are allotted time outside of the training program to investigate the synthetic learning environment, this could enhance their intrinsic motivation and encourage allocation of more cognitive resources towards schema acquisition and schema automation. Furthermore, research also shows that increased levels of intrinsic motivation generally result in a desire to spend more time working on the task (Deci & Ryan, 1987), which may lead to increased levels of learning. An additional factor that is critical to user satisfaction and use of synthetic learning environments is the instructor attitudes toward the training program (Sun, Tsai, Finger, Chen, & Yeh, 2008). Training instructors on the use of the simulation system and allowing time for them to explore the environment as well may stimulate excitement for pursuing new experiences, and could translate to the users and encourage motivation. Researchers of this study reported that the instructors often remarked that they wished they were able to use the program outside of training simulations; it would have been interesting to see the difference, if any, in attitudes towards the training programs.

It is also noteworthy that the participating soldiers were assessed on their learning the day following the training program in the synthetic learning environments, which may not have been representative of retention. The goal of training, regardless of the industry, is to not only achieve maximal training transfer, but also to ensure that these skills are retained for continued overall improvement on the job. The “training transfer” problem, in which Saks (2002) reported a meager 34% retention rate, was after a year of participation in the training program; comparatively, the retention rate was 62% immediately following the training program. Therefore, evaluating performance the day after a training program may not fully demonstrate the amount of training transfer that occurs as a result, and future research on this topic should endeavor to measure performance in the following months or years after the training program. This will allow researchers to properly identify a problem, if one does exist, diagnose potential sources, and advise strategies for targeted remedies. For instance, skill decay, which according to

Arthur, Bennett, Stanush, and McNelly (1998) “refers to the loss or decay of trained or acquired skills (or knowledge)” (p. 58), could be mitigated by follow-up training sessions after certain periods of time.

The use of VBS3 was called into question by several researchers not only for the lack of user-friendliness and source of high cognitive load, but also because of its cost and restrictive nature. Because VBS3 is a military simulation and training program that was based off of the combat game ARMA and is commercially owned, researchers were only able to alter the design of the training programs to a certain extent, and were not permitted full control over programming the interface and system. As such, there was no simple solution to the proposed lack of user interface as a cause for cognitive load. MOSES, however, is open source and therefore freely available and fully modifiable, which is ideal for the purposes of creating training programs unique to each training situation and industry.

Although this study did not achieve results in support of the proposed hypotheses, it is also important to acknowledge an absence of group differences between VBS3 and MOSES on workload and performance. This bodes well for further development in training through the use of MOSES, as it offers benefits such as cost reductions and ease in programmability. It could be interpreted that MOSES does not impede learning and training transfer, and, if this system continues to be an interest of the Army, should be further investigated. As advocated by experts in the training industry, it is always important to ensure that learners are not hindered by the system used to deliver the learning material; therefore, this study provides support for the shift from the use of VBS3 to MOSES for military training. Furthermore, the results of this study are not exclusive to the military field; other industries interested in incorporating technology into training and learning programs may also benefit and should take full advantage of the affordances of MOSES.

Limitations

The nature of field studies as opposed to laboratory experiments was the primary limitation for this effort due to lack of standardization and control. Specifically, the rubrics used to measure performance was an updated version based on a training manual that had already been in use and therefore had not been validated for research purposes. As discussed, instructions were minimal; while the instructors received directions on filling out the forms, there was no way to establish inter-rater reliabiltiy due to how training operations were run. Consequently, there was no way to ensure systematic ratings were being made, which may have contributed to inaccurate performance ratings and threatened the statistical conclusion validity of the study.

It was also difficult to ensure standardization between every group, as it was infeasible to require each training session to be conducted in the same classroom due to probable scheduling issues and that the Basic Leadership Course had a timeline independent of the researchers'. This meant that the training programs may have been conducted with different layouts (e.g. classroom configuration with rows facing the front, and another with rows facing each other) which may have interfered with group interactions. Instructors were also inconsistent across all groups and therefore the squad group leader's relationship with one group may have been different with another, or varying persoanlities between instructors may have influenced the dynamics and motivation of the groups differently.

All limitations mentioned were possibly a result of researchers trying to remain as noninvasive as possible, as per the agreements with the U.S. Army Research Laboratory and course instructors. As sessions were being run, researchers were not allowed to give any additional instructions or comments, or dictate how training sessions were implemented aside from the different training conditions. Furthermore, this study had a statistically small sample size, which may have partially explained the nonsignificant results; however, it was a large sample size for a field study conducted with the military. Regardless, field studies hold significance as they allow for research in contexts outside of a controlled laboratory and therefore contribute alternative perspectives on research variables and outcomes.

Areas of Future Research

The increasing incorporation of technology into training programs is not a novel idea, yet is something that can and should be constantly improved upon. This study highlights the importance of ensuring that future studies are designed to restrict the limitations that affect the strength of the conclusions that can be drawn. This study also warrants further investigation in identifying possible influences of cognitive load on performance measures in both a field and laboratory setting. As the shift towards implementing technology in our daily lives continues, there are several factors to consider.

Bloom’s Taxonomy (1956) proposes that there are different processes of cognition in learning, and is a concept that should be acknowledged when researching and developing training programs. The introduction of cognitive ability raises the question of if a synthetic learning environment is more conducive for a specific type of activity, whether the intention is targeted towards a learner *remembering* or *analyzing* the material, for example. Do learners benefit more from training programs if it simply teaches them to memorize a topic, or is it more helpful if the participant does a task that requires them to evaluate their situation? Do different objectives impose different cognitive loads, and are there situations that offer more opportunities for germane cognitive load processing? Such investigation would allow for the maximum return on investment as it will inform options for modifying training programs.

The purpose for this study was to continue the endeavor to maximize training through the use of technology, and ideally identify factors that influence training effectiveness. As the training industry enters a new era, future research ensuring that training effectiveness is optimized should remain a constant theme in both academic and practical fields.

Acknowledgments

This research was sponsored by the Army Research Laboratory and accomplished under Cooperative Agreement Number W911NF-14-2-0012. The views and conclusions contained in this document are those of the author's and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

APPENDIX A:
INFORMED CONSENT FORM

Understanding Virtual World Training Effectiveness – Phase 3

Informed Consent

Principal Investigators: Brian Goldiez, Ph.D. and Douglas Maxwell, Ph.D.

Sponsor: Army Research Laboratory - Human Research and Engineering Directorate Simulation and Training Technology Center (ARL HRED STTC)

Investigational Site: Florida National Guard Camp Blanding Joint Training Center located at 5629 State Road 16 West Bldg 2300 Avenue A, Starke, FL 32091

Introduction: Researchers at the University of Central Florida (UCF) study many topics. To do this we need the help of people who agree to take part in research studies. You are being invited to take part in a research study which will include up to 162 Soldiers. You must be 18 to 40 years of age, have U.S. citizenship, have normal or corrected to normal vision, and must not be colorblind to be included in the research study. Individuals who have a prior history of seizures cannot take part in this study. Your participation is voluntary and you may withdraw at any time. The people conducting this research are Dr. Brian Goldiez from the Institute for Simulation and Training at UCF and Douglas Maxwell from ARL HRED STTC.

What you should know about a research study:

- Someone will explain this research study to you.
- A research study is something you volunteer for.
- Whether or not you take part is up to you.
- You should take part in this study only because you want to.
- You can choose not to take part in the research study.
- You can agree to take part now and later change your mind.
- Whatever you decide it will not be held against you.
- Feel free to ask all the questions you want before you decide.

Purpose of the research study: The purpose of this research is to understand the application of emerging VW technologies within the typical training cycle (e.g., classroom, simulation, live). Specifically, this research focuses on virtual training technologies and strategies for collective training tasks associated with patrolling urban and rural environments.

What you will be asked to do in the study: You will be asked to complete a demographics survey about your age, sex, and your educational, military, computer-use, game-use, and simulation-based training background and experience. You may be asked to complete a color vision test. You will view training content for a patrolling urban and/or rural environments using traditional methods or within a Virtual World (VW). Following the brief, you may complete a series of training scenarios within a VW followed by a series of surveys related to your perceived level of usefulness, workload, presence, and engagement experienced during the training. After the training, you will receive a post-training brief and complete a Situational Training Exercise (STX).

Location: Florida National Guard Camp Blanding Joint Training Center located at 5629 State Road 16 West Bldg 2300 Avenue A, Starke, FL 32091

Time required: We expect that you will be in this research study for no more than nine hours. The maximum duration allotted for this experiment is twelve hours.

Audio or video taping: You will be video and audio taped during this study and you may be photographed during this study. Video and audio recordings and photographs will be considered privileged and held in confidence. If you do not want to be video or audio taped or photographed, then you may not be able to participate in the study. All video and audio recordings and photographs will be kept on a secure, password protected computer and/or in a locked cabinet at the Principal Investigators' work site at the Institute for Simulation and Training. If videotaping and photographing is approved, the recording and images will be erased or destroyed after all data is coded and the minimum time for storage of collected data has passed. **Please indicate below if you will allow us to take video and audio recordings and/or photographs during your experimental session.**

Do you give us your consent to be photographed, video recorded, and audio recorded during this study? *(Check one and initial)*

☐ Yes ☐ No Initial _____

Funding for this study: This research study is being paid for by ARL HRED STTC.

Risks: This study should offer minimal risks to your health and well-being. You can choose to withdraw from the experiment at any time, or request a break at any time. If you are a participant who receives the VW training experience, then you will complete a Simulator Sickness Questionnaire to monitor you for symptoms associated with simulator sickness (e.g., nausea, disorientation, visual disruptions). If you experience any of the symptoms mentioned, please tell the researcher and remain seated until the symptoms disappear.

Benefits: You will receive no benefits from participating in the experiment, other than gaining an increased knowledge and ability for conducting a STX mission, and the personal satisfaction of supporting the Army's research in developing improvements in Soldier training methods.

Compensation or payment: There is no compensation, other payment, or extra credit provided to you for taking part in this study.

Confidentiality: Your participation in this research is confidential. We will limit your personal data collected in this study to people who have a need to review this information. The data will be stored and secured in the offices of the principal investigator in a locked cabinet. The data, without any identifying information, will be transferred to a password-protected computer for data analysis. This consent form will be retained by the principal investigator for a minimum of three years. If the results of the experiment are published or presented to anyone, no personally identifiable information will be shared. The research staff will protect your data from disclosure to people not connected with the study. We cannot promise complete secrecy. Organizations that may inspect and copy your information include the UCF Institutional Review Board (IRB) and other representatives of UCF. In addition, since this research is sponsored by the Department of Defense and the U.S. Army, the Army Human Research Protections Office is eligible to review the research records.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, or think the research has hurt you in some way, you may contact Dr. Crystal Maraj at cmaraj@ist.ucf.edu or 407-882-2116.

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

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

Withdrawing from the study: Your decision to participate in this study is voluntary. You can stop at any time. You do not have to answer any questions you do not want to answer. Refusal to take part in or withdrawal from this study will involve no penalty or loss of benefits you would receive by staying in it. If you decide to leave the study, inform one of the research personnel so that you may be dismissed. The investigators or the sponsor can remove you from the research study without your approval or end the research study early. Possible reasons for removal or ending the study early include unanticipated disruptions or difficulties during the experimental sessions. We will tell you about any new information that may affect your health, welfare, or choice to stay in the research.

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

DO NOT SIGN THIS FORM AFTER THE IRB EXPIRATION DATE BELOW

Printed name of participant

Signature of participant

Date

Printed name of person obtaining consent

Signature of person obtaining consent

Date

APPENDIX B:
PERFORMANCE MEASURES

Squad Number: ______________________________

Evaluator: _______________________________________

1. Evaluation: please rate the squad's performance performing the following collective tasks during the assigned mission:

Task: React to Indirect Fire while Dismounted	**1 Needs Improvement**	**2 Adequate**	**3 Successful**	**4 Excels**
1. Shout "Incoming!" in a loud, recognizable voice.				
2. React to the instructions of your leader by listening and looking for guidance.				
3. Seek the nearest appropriate cover.				
4. Assess your situation.				
5. Report your situation to your leader.				
6. Continue the mission.				
Task: React to an IED	**1 Needs Improvement**	**2 Adequate**	**3 Successful**	**4 Excels**
1. Performed 0-5-25m checks.				
2. Establish 360-degree security.				
3. Employ tactical combat casualty care measures.				
4. Evacuate any casualties.				
5. Submit “Situation Report” to higher HQ’s.				
6. Continued mission IAW Higher HQ guidance.				
Task: React to Near Ambush	**1 Needs Improvement**	**2 Adequate**	**3 Successful**	**4 Excels**
1. Return fire immediately.				
2. Soldiers in the kill zone assault through the ambush using fire and movement.				
3. Soldiers not in the kill zone identify the enemy location, place "well-aimed" suppressive fire on the enemy's position and shift fire as Soldiers assault the objective.				
4. Soldiers assault through and destroy the enemy position.				
5. The unit leader reports the contact to higher headquarters.				
Task: React to Far Ambush	**1 Needs Improvement**	**2 Adequate**	**3 Successful**	**4 Excels**
1. Soldiers receiving fire immediately return fire, seek cover, establish a support by fire and suppress the enemy position(s).				

2. Soldiers not receiving fire move along a covered and concealed route to the enemies flank in order to assault the enemy position.				
3. The unit leader or Forward Observer calls for and adjusts indirect fires. On order, lifts or shifts fires to isolate the enemy position or to attack them with indirect fires as they retreat.				
4. Soldiers in the kill zone shift suppressive fires as the assaulting Soldiers fight through and destroy the enemy.				

2. **Legend**: please use the following rating definitions when evaluating the squad's performance:

Rating	**Definition**
Needs Improvement	Squad did not achieve this performance measure.
Adequate	Squad marginally achieved this performance measure.
Successful	Squad suitably achieved this performance measure.
Excels	Squad favorably achieved this performance measure.

APPENDIX C:
NASA-TLX QUESTIONNAIRE

Please rate your overall impression of demands imposed on you during the exercise.

1. Mental Demand: How much mental and perceptual activity was required (e.g., thinking, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

LOW |-----|----|----|----|----|----|----|----|----| HIGH
1 2 3 4 5 6 7 8 9 10

2. Physical Demand: How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

LOW |-----|----|----|----|----|----|----|----|----| HIGH
1 2 3 4 5 6 7 8 9 10

3. Temporal Demand: How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

LOW |-----|----|----|----|----|----|----|----|----| HIGH
1 2 3 4 5 6 7 8 9 10

4. Level of Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

LOW |-----|----|----|----|----|----|----|----|----| HIGH
1 2 3 4 5 6 7 8 9 10

5. Level of Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

LOW |-----|----|----|----|----|----|----|----|----| HIGH
1 2 3 4 5 6 7 8 9 10

6. Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

LOW |-----|----|----|----|----|----|----|----|----| HIGH
1 2 3 4 5 6 7 8 9 10

APPENDIX D:
DEMOGRAPHICS QUESTIONNAIRE

Demographics Questionnaire

Participant ID (ROSTER): _________ Date___________

1. General Information

Age (yrs): _____ Gender: _____M _____F

Do you have corrected vision? ___None ___Glasses ___Contact Lenses

Do you have any type of color blindness? ___Yes ___No

2. Military Experience

a. How many years have you been in the military? ________ Current rank ____________

b. What is your MOS? ________________

c. Have you served as a Team Leader? ____________ If yes, how long? ___________

d. Have you served as a Squad Leader? ____________ If yes, how long? ___________

e. Please list all combat deployments (Iraq, Afghanistan, etc.) and the length (Years / Months) of each.

Location	Time
________________________________	_____________________
________________________________	_____________________
________________________________	_____________________

Do you have training experience reacting to indirect fire while dismounted? ___Yes ___No

Do you have training experience in reacting to indirect fire while dismounted that used simulation or virtual reality? ___Yes ___No

If yes, what type and purpose?

Type	Purpose
_________________	____
_________________	____

Do you have training experience reacting to direct fire contact? ___Yes ___No

Do you have training experience in reacting to direct fire contact that used simulation or virtual reality? ___Yes ___No

If yes, what type and purpose?

Type	Purpose
___________________	____
___________________	____

Do you have training experience conducting a squad attack? ___Yes ___No

Do you have training experience in conducting a squad attack that used simulation or virtual reality? ___Yes ___No

If yes, what type and purpose?

Type	Purpose
___________________	____
___________________	____

3. Educational Data

a. What is your highest level of education received? Select one.

____ GED
____ High School
____ Some College
____ Bachelor’s Degree
____ M.S/M.A
____ Ph.D. or other doctorate
____ Other: _______________________

b. If applicable, what subject is your degree in (for example, Criminal Justice)?
____

4. Computer Experience

a. How long have you been using a computer?

____ Less than 1 year
____ 1-3 years
____ 4-6 years
____ 7-10 years
____ 10 years or more

b. How often do you use a computer?

____ Daily (please circle one): Over 2 hrs/day 1-2 hrs/day Less than 1 hr/day
____ Weekly
____ Monthly
____ Once or twice a year
____ Never

c. How often do you play video games?

____ Daily (please circle one): Over 2 hrs/day 1-2 hrs/day Less than 1 hr/day
____ Weekly
____ Monthly
____ Once or twice a year
____ Never

d. How often do you use a virtual world, such as Second Life?

____ Daily (please circle one): Over 2 hrs/day 1-2 hrs/day Less than 1 hr/day
____ Weekly
____ Monthly
____ Once or twice a year
____ Never

APPENDIX E: UCF IRB APPROVAL LETTERS

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

Approval of Human Research

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

To: **Stephanie Jane Lackey** and Co-PI: **Douglas B. Maxwell**

Date: **March 13, 2015**

Dear Researcher:

On 3/13/2015, the IRB approved the following human participant research until 03/12/2016 inclusive:

Type of Review: UCF Initial Review Submission Form
Expedited Review Category #6 and 7
Project Title: Understanding Virtual World Training Effectiveness – Phase 3
Investigator: Stephanie Jane Lackey
IRB Number: SBE-15-11077
Funding Agency:
Grant Title:
Research ID: 1052585

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 03/12/2016, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

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

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

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:

Joanne Muratori

Signature applied by Joanne Muratori on 03/13/2015 04:07:56 PM EDT

IRB manager

University of Central Florida

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

Approval of Human Research

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

To: **Crystal Maraj** and Co-PI: **Douglas B. Maxwell**

Date: **February 16, 2016**

Dear Researcher:

On 02/16/2016, the IRB approved the following minor modifications to human participant research until 02/04/2017 inclusive:

Type of Review: IRB Addendum and Modification Request Form
Expedited Review Category #6 and 7
Modification Type: The PI on the study has been changed from Dr. Brian Goldiez to Dr. Crystal Maraj. New research associates, William Aubrey, Samantha Napier, and Jonathan Hurter have been added. The following KSP have been removed from the study: Karla Badillo-Urquiola, Eric Ortiz, and Veronica Prisco. A revised Informed Consent document has been approved for use.
Project Title: Understanding Virtual World Training Effectiveness – Phase 3
Investigator: Crystal Maraj
IRB Number: SBE-15-11077
Funding Agency:
Grant Title:
Research ID: 1052585

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 02/04/2017, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

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

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

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:

Joanne Muratori

Signature applied by Joanne Muratori on 02/16/2016 12:22:46 PM EST

IRB Manager

APPENDIX F: ARL IRB APPROVAL LETTERS

DEPARTMENT OF DEFENSE
UNITED STATES OF AMERICA

DEPARTMENT OF THE ARMY
US ARMY RESEARCH, DEVELOPMENT AND ENGINEERING COMMAND
ARMY RESEARCH LABORATORY
BUILDING 459
ABERDEEN PROVING GROUND MD 21005-5425

REPLY TO
ATTENTION O

RDRL-HRS 10 February 2015

MEMORANDUM FOR: Douglas Maxwell, STTC, HRED, ARL, Orlando, FL
FROM: Theresa M. Straut, Human Protection Administrator,
Army Research Laboratory IRB
SUBJECT: Approval of Research Study, ARL 15-010

PROJECT TITLE: Understanding Virtual World Training Effectiveness – Phase 3
SUBMISSION TYPE: Initial Protocol
REVIEW TYPE: Expedited Review
APPROVAL PERIOD: 10 February 2015 to 9 February 2016

The purpose of this memorandum is to notify you that the research project identified above has been approved by the ARL Institutional Review Board (IRB) by expedited review under category 7 on 10 February 2015.

The project documents were reviewed:

- Protocol dated 020615
- Consent dated 020615
- Demographic Questionnaire
- DSSQ Short Form
- Engagement Measure MOSES
- Flow State Short Scale
- NASA-Task Load Index
- Presence Questionnaire
- Simulator Sickness Questionnaire
- Technology Acceptance Measure

On 9 Feb 2015 a revised protocol, dated 02092015, was submitted to address clarifications requested by the IRB reviewer.

An IRB approved informed consent form was not generated by the ARL IRB. The ARL IRB assumes that the University of Central Florida (UCF) will be generating an IRB approved version of the consent. If any alteration to the content of the consent form is made, the ARL IRB will need to re-review the document.

As principal investigator, you are responsible for ensuring that the study is conducted in accordance with the final version of your protocol. You cannot delegate your supervisory responsibility to anyone else associated with the project. If you leave the project a new principal investigator should be designated for the research. Designation of a new principal investigator should be reported to the IRB.

In addition, you must report the following to the IRB:

- You must report changes in research personnel, including the principal investigator, involved in the study.
- You must report changes in the research procedures before they are initiated. You can report minor changes by completing the ARL amendment form.
- You may make changes in research procedures implemented to eliminate immediate hazards to the subjects, but they must be reported within 10 days of their implementation on the amendment form.
- You must report completion or discontinuation of your study by submitting a completion or discontinuation report to the IRB.
- You must report plans to continue your study beyond the expiration date before you attain that date, by submission of a continuing review form 30 days before the expiration date.
- You must promptly report any injury or Unanticipated Problems Involving Risks to Participants or Others (UPIRTSO) to the IRB within 24 hours (via phone message, e-mail, or written report) of the incident. This should be followed by a full written report within 10 business days.

  A UPIRTSO is defined in DODI 3216.02, Glossary as “Any incident, experience, or outcome that meets ALL three of the following conditions:

  a. Is unexpected (in terms of nature, severity, or frequency) given the procedures described in the research protocol documents (e.g., the IRB-approved research protocol and informed consent document) and the characteristics of the human subject population being studied.
  b. Is related or possibly related to participation in the research. *Possibly related* means there is a reasonable likelihood that the incident, experience, or outcome may have been caused by the procedures involved in the research.
  c. Suggests that the research places human subjects or others at a greater risk of harm (including physical, psychological, economic, or social harm) than was previously known or recognized, even if no harm has actually occurred.

Good luck with your research.

STRAUT.THERES
A.M.1501857242
Digitally signed by STRAUT.THERESA.M.1501857242
DN: c=US, o=U.S. Government, ou=DoD, ou=PKI, ou=USA, cn=STRAUT.THERESA.M.1501857242
Date: 2015.02.10 14:37:40 -05'00'

Theresa M. Straut, CIP, RAC
Human Protection Administrator

DEPARTMENT OF DEFENSE
UNITED STATES OF AMERICA

DEPARTMENT OF THE ARMY
US ARMY RESEARCH, DEVELOPMENT AND ENGINEERING COMMAND
ARMY RESEARCH LABORATORY
BUILDING 459
ABERDEEN PROVING GROUND MD 21005-5425

REPLY TO
ATTENTION TO

RDRL-HR 19 Jul 2016

MEMORANDUM FOR: Douglas Maxwell, ARL-HRED, Orlando, FL

FROM: Theresa M. Straut, Human Protection Administrator, ARL, APG, MD

SUBJECT: Approval of Continuation of Research, ARL 15-074

Project Title: Aviation Combined Arms Tactical Trainer (AVCATT) Training Effectiveness Evaluation (TEE)
Submission Type: Continuing Review
Approval Period: 18 Jul 2016 to 17 Jul 2017

The purpose of this memo is to notify you that your continuing review was reviewed and approved by the IRB. Approval for your project expires on 17 Jul 2017, one year after the continuing review date, unless suspended or terminated earlier by action of the ARL Institutional Review Board (ARL IRB). When you complete your study, please submit a study closure report to the ARL IRB.

The following items were received and reviewed:

- Continuing Review Form

As principal investigator, you are responsible for ensuring that the study is conducted in accordance with the final version of your protocol. You cannot delegate your supervisory responsibility to anyone else associated with the project. If you leave the project a new principal investigator should be designated for the research. Designation of a new principal investigator should be reported to the ARL IRB.

In addition, you must report the following to the ARL IRB:

- You must report changes in research personnel, including the principal investigator, involved in the study.
- You must report changes in the research procedures before they are initiated. You can report minor changes by completing the ARL amendment form.

- You may make changes in research procedures implemented to eliminate immediate hazards to the subjects, but they must be reported within 10 days of their implementation on the amendment form.
- You must report completion or discontinuation of your study by submitting a completion or discontinuation report to the ARL IRB.
- You must report plans to continue your study beyond the expiration date before you attain that date, by submission of a continuing review form 30 days before the expiration date.
- You must promptly report any injury or Unanticipated Problems Involving Risks to Participants or Others (UPIRTSO) to the ARL IRB within 24 hours (via phone message, e-mail, or written report) of the incident. This should be followed by a full written report within 10 business days.

A UPIRTSO is defined in DODI 3216.02, Glossary as “Any incident, experience, or outcome that meets ALL three of the following conditions:

- Is unexpected (in terms of nature, severity, or frequency) given the procedures described in the research protocol documents (e.g., the IRB-approved research protocol and informed consent document) and the characteristics of the human subject population being studied.
- Is related or possibly related to participation in the research. *Possibly related* means there is a reasonable likelihood that the incident, experience, or outcome may have been caused by the procedures involved in the research.
- Suggests that the research places human subjects or others at a greater risk of harm (including physical, psychological, economic, or social harm) than was previously known or recognized, even if no harm has actually occurred.

Should you have any questions or concerns, you can contact me at 410-278-5928 or at theresa.m.straut.civ@mail.mil.

STRAUT.THERES
A.M.150185724
2

Digitally signed by STRAUT.THERESA.M.1501857242
DN: c=US, o=U.S. Government, ou=DoD, ou=PKI, ou=USA, cn=STRAUT.THERESA.M.1501857242
Date: 2016.07.19 07:51:46 -04'00'

THERESA M. STRAUT, CIP, RAC
Human Protection Administrator, ARL

REFERENCES

Aguinis, H., & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology, 60*, 451-474.

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology, 42*(2), 331-342.

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*(2), 341-358.

Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review, 3*(4), 385-416.

Arthur, W., Bennett, W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88*(2), 234.

Arthur, W., Bennett, W., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*(1), 57-101.

Arthur, W., Tubré, T., Paul, D. S., & Edens, P. S. (2003). Teaching effectiveness: The relationship between reaction and learning evaluation criteria. *Educational Psychology, 23*(3), 275-285.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics, 60*(1), 47-57.

Ashenfelter, O., & Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics, 67*(4), 648-660.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation, 2*, 89-195.

Backlund, P., & Hendrix, M. (2013). Educational games - are they worth the effort? A literature survey of the effectiveness of serious games. *Games and virtual worlds for serious applications (VS-GAMES), 2013 5th international conference* (pp. 1-8). IEEE.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*(1), 63-105.

Bell, B. S., & Kozlowski, S. W. (2007). Advances in technology-based training. In S. Werner (Ed.), *Managing human resources in north america: Current issues and perspectives.* (pp. 27-42). Routledge.

Bellotti, F., Berta, R., & De Gloria, A. (2010). Designing effective serious games: Opportunities and challenges for research. *International Journal of Emerging Technologies in Learning., 5*.

Bloom, B. S., Krathwohl, D. R., & Masia, B. B. (1956). *Taxonomy of educational objectives: The classification of educational goals.* New York, NY: D. McKay.

Bonk, C. J., & Dennen, V. P. (2005). *Massive multiplayer online gaming: A research framework for military training and education (No. TECH-RPT-2005-1).* INDIANA UNIV AT BLOOMINGTON.

Bowers, C. A., Braun, C. C., & Morgan, B. B. (1997). Team workload: Its meaning and measurement. In M. Brannick, E. Salas, & C. Prince, *Team performance assessment and measurement: Theory, methods, and applications.* (pp. 85-108). Psychology Press.

Brandon Hall Group. (2015). *State of learning and development: The struggle with strategy.* Delray Beach, FL: Brandon Hall Group Research Team.

Broad, M. L., & Newstrom, J. W. (1992). *Transfer of training: Action-packed strategies to ensure high payoff from training investments.* Reading, MA: Corporate and Professional Publishing Group, Addison-Wesley Publishing Co.

Brogden, H. E., & Taylor, E. K. (1950). The dollar criterion: Applying the cost accounting concept to criterion construction. *Personnel Psychology, 3*(2), 133-154.

Burke, L. A., & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human Resource Development Review, 6*(3), 263-296.

Burke, M. J., & Day, R. R. (1986). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology, 71*, 232-245.

Cannon-Bowers, J., & Bowers, C. (2009). Synthetic learning environments: On developing a science of simulation, games, and virtual worlds for training. In S. Kozlowski, & E. Salas, *Learning, training, and development in organizations* (pp. 229-261). Taylor & Francis.

Card, D., & Sullivan, D. M. (1987). Measuring the effect of subsidized training programs on movements in and out of employment. *Econometrica, 56*(3), 497-530.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*(2), 233-246.

Cheng, E. W., & Ho, D. C. (2001). A review of transfer of training studies in the past decade. *Personnel Review, 30*(1), 102-118.

Collins, D. B., & Holton, E. F. (2004). The effectiveness of managerial leadership development programs: A meta-analysis of studies from 1982-2001. *Human Resource Development Quarterly, 15*(2), 217-248.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*(5), 678-707.

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105-134.

Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology, 53*(6), 1024.

Dickey, M. D. (2005). Three-dimensional virtual worlds and distance learning: Two case studies of Active Worlds as a medium for distance education. *British Journal of Educational Technology , 36*(3), 439-451.

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist, 63*(7), 591.

Farrington, J. (2011). From the research: Myths worth dispelling: Seriously, the game is up. *Performance Improvement Quarterly, 24*(2), 105-110.

Ford, J. K., & Weissbein, D. A. (1997). Transfer of training: An updated review and analysis. *Performance Improvement Quarterly, 10*(2), 22-41.

GAO. (2016). *Army training. Efforts to ajdust training requirements should consider the use of virtual training devices.* Retrieved from http://www.gao.gov/assets/680/679181.pdf

Georgenson, D. L. (1982). The problem of transfer calls for partnership. *Training and Development Journal*.

Gerjets, P., & Scheiter, K. (2003). Goal configurations and processing strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational Psychologist, 38*(1), 33-41.

Gerjets, P., Scheiter, K., & Cierniak, G. (2009). The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories. *Educational Psychology Review, 21*(1), 43-54.

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning, 29*(3), 207-219.

Greitzer, F. L., Kuchar, O. A., & Huston, K. (2007). Cognitive science implications for enhancing training effectiveness in a serious gaming context. *Journal on Educational Resources in Computing, 7*(3), 2.

Hadi, N. U., Abdullah, N., & Sentosa, I. (2016). Making sense of mediating analysis: A marketing perspective. *Review of Integrative Business and Economics Research, 5*(2), 62-76.

Haque, S., & Srinivasan, S. (2006). A meta-analysis of the training effectiveness of virtual reality surgical simulators. *IEEE Transactions on Information Technology in Biomedicine, 10*(1), 51-58.

Harman, E. A., Gutekunst, D. J., Frykman, P. N., Nindl, B. C., Alemany, J. A., Mello, R. P., & Sharp, M. A. (2008). Effects of two different eight-week training programs on military physical performance. *The Journal of Strength and Conditioning Research, 22*(2), 524-534.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati, *Human Mental Workload* (pp. 139-183). Amsterdam: North Holland Press.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford Publications.

Holton, E. F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly, 7*(1), 5-21.

Holton, E. F., & Baldwin, T. T. (2000). Making transfer happen: An action perspective on learning transfer systems. *Advances in Developing Human Resources, 8*(2), 1-6.

IBM. (2008). *Unlocking the DNA of the Adaptable Workforce, the IBM Global Human Capital Study.* Milwaukee, WI: IBM.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, interference and sensitivity analysis for causal mediation effects. *Statistical Science*, 51-71.

Industry Report (2005). *Training Magazine*. Retrieved from http://www.cedma-europe.org/newsletter%20articles/Training%20Magazine/Training%20Magazine%20Industry%20Report%202005%20(Dec%2005).pdf

Industry Report (2017). *Training Magazine.* Retrieved from https://trainingmag.com/trgmag-article/2017-training-industry-report

Jarmon, L., Traphagan, T., Mayrath, M., & Trivedi, A. (2009). Virtual world teaching, experiential learning, and assessment: An interdisciplinary communication course in Second Life. *Computers and Education, 53*(1), 169-182.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5*(5), 602-619.

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

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

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*(1), 1-17.

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*(1), 5-23.

Kalyuga, S., Chandler, P., & Sweller, P. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-372.

Kalyuga, S., Chandler, P., & Sweller, P. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*, 126-136.

Kanfer, R., & Ackerman, P. L. (1993). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*(4), 657-690.

Kaplan, R. M., & Pascoe, G. C. (1977). Humorous lectures and humorous examples: Some effects upon comprehension and retention. *Journal of Educational Psychology, 69*(1), 61.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology, 93*, 59-69.

Kirkley, S. E., Tomblin, S., & Kirkley, J. (2005). Instructional design authoring support for the development of serious games and mixed reality training. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).*

Kirkpatrick, D. L. (1959a). Techniques for evaluating training programs. *Journal of ASTD, 13,* 3-9.

Kirkpatrick, D. L. (1959b). Techniques for evaluating training programs: Part 2-Learning. *Journal of ASTD, 13*, 21-26.

Kirkpatrick, D. L. (1960a). Techniques for evaluating training programs: Part 3-Behavior. *Journal of ASTD, 14*, 13-18.

Kirkpatrick, D. L. (1960b). Techniques for evaluating training programs: Part 4-Results. *Journal of ASTD, 14*, 28-32.

Kirkpatrick, D. L. (1967). Evaluation of Training. In R. L. Craig, & L. R. Bittel (Eds.), *Training and development handbook* (pp. 87-112). New York: McGraw-Hill.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*, 1-10.

Kraiger, K. (2002). *Creating, implementing, and managing effective training and development: State-of-the-art lessons for practice.* Jossey-Bass.

Lackey, S. J., Salcedo, J., Matthews, G., & Maxwell, D. (2014). Virtual world room clearing: A study in training effectiveness. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, (pp. 1-11). Orlando.

Maxwell, D. (2015). Gauging training effectiveness of virtual environment simulation-based applications for an infantry soldier training task (Unpublished doctoral dissertation). Orlando, FL: University of Central Florida.

Maxwell, D., & Zheng, J. (2017). Large scale testing and evaluation of virtual environments for infantry soldier tasks comparing mental effort for live versus virtual training performance assessments. *MODSIM World*, (pp. 1-10).

Maxwell, D., Stevens, J., & Maraj, C. (2016). Alternate rubric for performance assessment of infantry soldier skills training. *International Conference on Augmented Cognition* (pp. 34-45). Cham: Springer.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology, 100*, 509-539.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.

Michalak, D. F. (1981). The neglected half of training. *Training and Development Journal, 35*(5), 22-28.

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999-2009). *Computers and Education, 56*(3), 769-780.

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

Mimeo. (2017). *Report: The state of learning and development in 2017.* Retrieved from https://www.mimeo.com/wp-content/uploads/Mimeo-InSync-State-of-LD-2017-Final-Ebook.pdf

Munro, A., Patrey, J., Biddle, E. S., & Carroll, M. (2015). Cognitive aspects of virtual environment design. In K. S. Hale, & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 391-410). Boca Raton: Taylor & Francis Group, LLC.

Neily, J., Mills, P. D., Young-Xu, Y., Carney, B. T., West, P., Berger, D. H., . . . Bagian, J. P. (2010). Association between implementation of a medical team training program and surgical mortality. *American Medical Association, 304*(15), 1693-1700.

Nelson, B., & Erlandson, B. (2008). Managing cognitive load in educational multi-user virtual environments: Reflection on design practice. *Educational Technology Research and Development, 56*, 619-641.

Noe, R. A. (1986). Trainees' attributes and attitudes: Neglected influences on training effectiveness. *Academy of Management Review, 11*(4), 736-749.

Paas, F. G., & Van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371.

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

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, 1-8.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1-8.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Reinerman-Jones, L., Goodwin, M. S., Goldiez, B. F., & Wismer, A. J. (2017). Evaluating game-based environments for army aviation and collective training. *MODSIM World*, (pp. 1-11).

Rodin, M., & Rodin, B. (1972). Student evaluations of teachers. *Science, 177*(4055), 1164-1166.

Saks, A. M. (2002). So what is a good transfer of training estimate? A reply to Fitzpatrick. *The Industrial-Organizational Psychologist, 39*(3), 29-30.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*(1), 471-499.

Salas, E., & Stagl, K. C. (2009). Design training systematically and follow the science of training. *Handbook of principles of organizational behavior: Indispensible knowledge for evidence-based management*, 59-84.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology, 8*(3), 197-208.

Salas, E., DiazGranados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008). Does team training improve team performance? A meta-analysis. *Human Factors, 50*(6), 903-933.

Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest, 13*(2), 74-101.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*(4), 469-508.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*(2), 489-528.

Sleight, D. (1993). A developmental history of training in the United States and Europe. Retrieved from https://msu.edu/~sleightd/trainhst.html

Sun, P. C., Tsai, R. J., Finger, G., Chen, Y. Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors inflluencing learner satisfaction. *Computers and Education, 50*(4), 1183-1202.

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

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295-312.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185-233.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). The redundancy effect. In *Cognitive Load Theory* (pp. 141-154). Springer, New York.

Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Tannenbaum, S. I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology, 43*(1), 399-441.

Tannenbaum, S. I., Cannon-Bowers, J. A., Salas, E., & Mathieu, J. E. (1993). *Factors that influence training effectiveness. A conceptual model and longitudinal analysis (No. NAVTRASYSCEN-TR-93-011).* NAVAL TRAINING SYSTEMS CENTER ORLANDO FL.

Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology, 90*(4), 692-709.

Vallejo Seco, G., Ato García, M., Fernández García, M., & Livacic Rojas, P. E. (2013). Multilevel bootstrap analysis with assumptions violated. *Psicothema, 25*(4).

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

Van Merriënboer, J. J., Jelsma, O., & Paas, F. G. (1992). Training for reflective expertise: A four-component instructional design model for complex cognitive skills. *Educational Technology Research and Development, 40*(2), 23-43.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research, 11*(4), 342-365.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34*(3), 229-243.

Vygotsky, L. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Wexley, K. N., & Baldwin, T. T. (1986). Posttraining strategies for facilitating positive transfer: An empirical exploration. *Academy of Management Journal, 29*(3), 503-520.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in Educational Theory and Practice, 93*, 27-30.

Woodworth, R. S., & Thorndike, E. L. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8*(3), 247.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 249.

Zyda, M. (2005). From visual simulation to virtual reality games. *Computer, 38*(9), 25-32.