2015

Gauging Training Effectiveness of Virtual Environment Simulation Based Applications for an Infantry Soldier Training Task

Douglas Maxwell
University of Central Florida

GAUGING TRAINING EFFECTIVENESS OF VIRTUAL ENVIRONMENT SIMULATION BASED APPLICATIONS FOR AN INFANTRY SOLDIER TRAINING TASK

by

DOUGLAS BRENT MAXWELL
B.S. Louisiana Tech University, 1998
M.S. Louisiana Tech University, 2001
M.S. University of Central Florida, 2014

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Modeling and Simulation
in the College of Graduate Studies
at the University of Central Florida
Orlando, Florida

Summer Term
2015

Major Professor: Peter Kincaid

ABSTRACT

The U.S. Army Training and Doctrine Command's Army Learning Concept 2015 and Army Training Concept 2025 are documents that discuss the need for an adaptive soldier learning model with a flexible training delivery methodology. The U.S. Army has been investing in serious gaming technology for the past two decades as a cost effective means to teach tactics and strategy. Today, the U.S. Army is seeking to expand its application of virtual environment training to areas such as cultural awareness and human network analysis for the infantry soldier. These new expanded applications will require a higher level of non-determinant behavior inside the virtual environment.

To meet more of the training needs of the war fighter, the U.S. Army is looking beyond first person perspective games to the cooperative and social gaming experience offered by the MMOG (Massively Multiplayer Online Game) and the VWT (Virtual World Technology). Altogether, these classes of games have the potential to teach leadership skills, social acclimation skills, cultural awareness and practice skills, and critical thinking skills for problem solving in a cost effective manner. Unfortunately, even today there is a paucity of scientific research to support whether this potential may be realized or not (Pennell, 2003; Whitney, Temby, & Stephens, 2013).

A literature review was performed which covers current concepts in the usage of virtual environments for military individual and team training in the U.S. Army infantry soldier domains. There are many variables involved with the lifecycle of the virtual training activity including the acquisition, information assurance and cyber security, deployment, proper employment, content development and maintenance, and retirement. This discussion goes beyond the traditional topics

of graphics and game engine technology and delves deeper into concepts of the importance of proper usage of the environments by the trainees.

This dissertation is composed of three studies with two subject pools: experienced soldiers and novice soldiers. The participants in the studies were randomly assigned to one of two training conditions. The training conditions were either a traditional slide-show in a classroom or a virtual environment based training system. The participants were then provided with training for a room clearing tasks in each of the conditions. The independent variables are training condition and soldier condition. The dependent variables are individual performance, team performance, stress questionnaire scores, and workload questionnaire scores.

A number of relationships are explored in this dissertation. The first objective of these studies is to attempt to identify any effect the training conditions have on either individual performance or team performance. Lastly, these studies attempt to identify if there is any difference the training conditions have on novice versus experienced subjects' performance during a live assessment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS (or) ABBREVIATIONS

ALC2015	Army Learning Concept 2015
AMRDEC	Aviation and Missile Research Development and Engineering Center
AMSAA	Army Materiel Systems Analysis Activity
ARL	U.S. Army Research Laboratory
ATC2025	Army Training Concept 2025
CA	Cooperative Agreement
COTS	Commercially Available/Off the Shelf
DARPA	Defense Advanced Research Programs Agency
DIS	Distributed Interactive Simulation
DSSQ	Dundee Stress State Questionnaire
FLARNG	Florida Army National Guard
GBVE	Game Based Virtual Environment
GOTS	Government Available/Off the Shelf
HLA	High Level Architecture
IED	Improvised Explosive Device
JIEDDO	Joint Improvised Explosive Devices Defeat Organization
LVC	Live, Virtual, Constructive Simulations
M&S	Modeling and Simulation
MMOG	Massively Multiplayer Online Game
MOS	Military Occupational Specialties
MOSES	Military Open Simulator Enterprise Strategy
OC	Observer Controller
OE	Operational Environment

PEO-STRI	Program Executive Office – Simulation, Training, & Instrumentation
PKAD	Physical Knowledge Acquisition Data
RTI	Regional Training Institute
ROTC	Reserve Officer Training Corps
SBT	Simulation Based Training
SME	Subject Matter Expert
STM	Science and Technology Manager
STTC	U.S. Army Simulation and Training Technology Center
TRADOC	U.S. Army Training and Doctrine Command
VWT	Virtual World Technology
WLC	Warrior Leader Course

CHAPTER ONE: INTRODUCTION

Chapter One Summary

Chapter one provides the motivation and argument for this research as well as a description of gaps and a detailed description of the challenges facing the training community in the U.S. Army in the current constrained fiscal climate. Sequestration is forcing reductions in U.S. military training budgets, however ongoing global instability requires the U.S. Army to retain capability.

Context

The United States military is investing significant resources into the use of virtual environments for war fighter training applications. In the past, success with flight simulators and vehicle simulators have shown tremendous savings in lives, money, and time (Rushmer, 2006; Keh, Wang, & Wai, 2008). These successes were driven by straightforward applications of the technology and could be considered “low hanging fruit”.

The quality of today’s United States infantry soldiers is widely accepted as among the best in the world. Examination of tactical engagements in Iraq and Afghanistan reveal enemy combatants running and shooting wildly while American soldiers maintain tight formations and carry their rifles with fingers outside the trigger wells (Scales, 2013). This is a testament to superior training, which the U.S. soldiers appreciate more than pay and benefits as they recognize this preparation is the best life insurance.

In previous eras, this proficiency was earned through experience at a very high cost. For example, in Vietnam around two-thirds of small unit infantry soldier casualties happened within the first two months of deployment. This is due to the accelerated pace at which the training system had to produce soldiers, too quickly to prepare them for the operating environment and tasks of close-combat killing (Scales, 2013). The lesson learned by senior leadership was that infantry soldiers required much more rigorous training before exposing them to actual combat conditions.

Rigorous preparations drive the requirements for realistic training, especially as close combat activities become more vital as the American military pivots in the Middle East. In 2012, the United States Military spent $172 billion in support of training and readiness out of a total defense budget of $531 billion (Office of Management and Budget, 2014). Reductions in military budgets have been enforced due to the sequestration activities by congress, meaning alternatives to the currently costly live combat training simulations must be found (Osborn, 2015).

Today, the leadership in the military training community is seeking to expand the application of virtual environment infantry soldier training to areas such as cultural awareness and human network analysis. These new expanded applications require a higher level of non-determinant behavior inside the virtual environment. The Army Learning Concept 2015 (ALC2015) outlines the need for new learning models to produce soldiers with the ability to be adaptable and utilize critical thinking skills. The ALC2015 discusses the need to focus on individual soldier performance and leadership learning in all areas from initial military training to professional military education and functional coursework. Soldiers must be continuously adaptive with a flexible training delivery infrastructure that promotes learning throughout the entire career of the soldier (Morton, Lucious (Department of the Army, HQ Deputy Chief of Staff, 2011)).

The ALC2015 makes a number of assumptions to arrive at its discussed needs. It is assumed the U.S. Army will operate under a continued era of uncertainty and persistent conflict. The U.S. Army will continue to be confronted with unexpected challenges from adversaries that are adaptive. This necessitates the rapid development of training and education as well as changes in doctrine to match the adaptations. The ALC2015 also recognizes that learning is a lifelong endeavor, performed during the entire course of a career and not limited in duration or location. The learning must be accessible from anywhere and at the point of need. Conversely, the soldier must have an opportunity to contribute back to the body of knowledge.

The ALC2015 calls for an immediate shift from the current training models by making changes in the way the courses are presented. It calls for classrooms to be converted into collaboration spaces led by facilitators. The goal is to engage soldiers and students and encourage them to think and understand what is being presented through interaction and discourse. Also, the ALC2015 calls for the reduction or elimination of slide based presentations and lectures. Blended learning principles should be adopted that incorporate constructive simulations with virtual environments, gaming technology, or web based delivery.

The U.S. Army Training Concept 2012 – 2020 (ATC2020) TRADOC Pam 525-8-3 (Morton, Lucious (Department of the Army, HQ, Colonel, GS, Deputy Chief of Staff, 2011)) expands upon the concepts outlined in the ALC2015 and calls for more realistic training opportunities and experiences. Technological innovation is required to realize the need to replicate the ambiguities and uncertainty of actual missions. Future training systems need more non-determinant environments that allow for soldiers and leaders to exercise free will in decision making to accomplish a mission.

A key goal described in the ATC2020 is the concept of foundational home station training for domain relevant training. This means soldiers must be provided with the tools and technology to allow for the portrayal of the operational environment (OE) to a sufficient fidelity on demand. Additionally, this new distributed training concept must allow for training not just at the small unit level but also cooperative training through the higher echelons as well.

ATC2025 calls for technology development to enable collective training for soldiers and leaders geographically dispersed. This requirement is a reaction to constrained budgets that will prevent the relocation of personnel to physical training locations in certain situations, however the training activity will still need to be performed to maintain proficiency. An example of expanded skill training cited in the concept document is cultural awareness and improvised explosive device identification training. Current training systems such as the First Person Cultural Trainer (FPCT) sponsored by the U.S. Army's Training and Doctrine Command and the I-Game sponsored by the Joint Improvised Explosive Device Defeat Organization (JIEDDO) illustrate significant initial investments by the U.S. military in this domain (IPKeys, 2014; Zielke, Zakhidov, Hardee, & Kaiser, 2014).

Motivation for Research

The United States Army has invested significant funding dedicated to the use of virtual environments for training infantry soldier skills. There is a pervasive attitude in the acquisition community that a simulation based training system's graphics quality are the strongest indicators of utility and training quality. Very little data exists to quantify the return on investment provided by these training systems. There is also a lack of formal methodologies for the identification of where in the training cycle these technologies belong as well as which training tasks they should be applied. The United States Government Accountability Office issued a report in August of

2013 which calls for better assessment of performance and accounting of costs to properly assess simulation based training systems throughout the U.S Army and Marine Corps (Pickup, 2013). It is the author's personal experience in the U.S. Army simulation based training research, acquisition decisions are often made with a bias towards the visual fidelity of the product rather than a total assessment of the training system's functionality. Since there is a lack of empirical data for a particular task's training effectiveness using virtual environments (Haque & Srinivasan, 2006), there is little guidance for the program manager's decision making processes. This leaves the requirements generation team and the acquisition process to attempt to replicate the training provided by traditional means in a virtual environment. There is too much leeway in the interpretation of this replication and final decisions are not based on established scientific basis but the desires of the target user community.

To further complicate matters, the lack of formal requirements and performance measurement methodologies has led to a fracturing of the training space within the U.S. military that utilizes game based virtual environments. Although there is a game based virtual environment for training program of record called Virtual Battle Spaces, it is not considered sufficient for the specialized training needs of some organizations. Pockets of innovation and product development has taken place in recent years, resulting in numerous training systems specializing in different utilizations. Table 1 shows a sampling of the trade space of the available programs, their intended utility, and sponsors. There is overlap in a number of these programs, especially in the systems utilizing the same base game engine. For example the FPCT, America's Army, I-GAME, and EDGE are all based on a commercial game engine called Unreal (EPIC Games, 2014a). It is the author's opinion this method of virtual simulation development is unnecessarily duplicative and the "re-licensing" of the same engine at great expense is wasteful.

Table 1. Examples of U.S. Army SBT Systems for Ground Skills Training

Program Name	Utility	Sponsor
America's Army	Game based soldier simulation intended for recruitment.	Army Game Studio Redstone Arsenal AMRDEC
Close Combat Tactical Trainer (CCTT)	Simulation composed of three modules with full-crew simulators, mock combat outposts, and infantry soldier arms training.	PEO-STRI
DARWARS *Ambush*	Game based simulation for convoy contingency training.	DARPA
Enhanced Dynamic Geo-social Environment (EDGE)	Prototype game based emergency response and cultural training simulation.	ARL/STTC
I-GAME	Game based virtual environment for scenario building and mission rehearsal training with focus on improvised explosive devices	JIEDDO
First Person Cultural Trainer (FPCT)	Game based simulation for cultural interaction between US soldiers and various foreign populations.	TRADOC Intelligence Support Activity
Military Open Simulator Enterprise Strategy (MOSES)	Prototype: Persistent virtual world for experimental use in infantry soldier training effectiveness evaluation.	ARL/STTC
Virtual Battle Spaces (VBS2/VBS3)	Game based virtual environment for mission rehearsal and scenario training.	PEO-STRI

This dissertation describes the first phase of a three-phase research project conducted through Cooperative Agreements (CA) #W911NF-14-0012 and #W911NF-15-0004 between the U.S. Army Research Laboratory and the University of Central Florida. These CAs were created to facilitate the investigation of training effectiveness of operationally relevant tasks in a virtual environment as compared to traditional classroom and live means. The desired outcome of this work is to establish a methodology for quantitatively defining the training effectiveness differences between traditional and virtual methods and acquiring data through field experimentation to exercise the methodology.

The original intent for this dissertation was to design and execute experimentation to compare effectiveness of various virtual environment technologies for select ground skills task training. It became apparent early in the literature search and interview of numerous subject matter experts that this examination is premature. The U.S. Army has not yet established concrete assessment of performance of the use of virtual environments. A "crawl, walk, run" methodology is performed to the data collection methodology for the larger Cooperative Agreement, with the crawl phase being the assessment of a training task conducted with a large number of soldiers using a generic virtual environment as a comparison to the traditional condition. It is only after the establishment of the differences between traditional and virtual training can we then look at the nuanced differences between the various virtual training products.

Description of Gap

The U.S. military has a significant investment in the research and development of virtual environments for training. Further, there is a significant investment in the fielding and maintenance of virtual environments. There is no dispute that virtual training costs less than live training exercises, but the services are not currently able to quantify the cost differences. Additionally, the value of simulation for training is subjectively assessed by leaders based on performance standards and the assessor's expertise (Insinna, 2013). Currently, training is developed in virtual environments by translating the instruction from the live and classroom environments to the simulators. There are inherent differences between the virtual environments and the live/classroom training techniques that must be accounted for.

The following literature review covers current concepts in the usage of virtual environments for military individual and team training in the U.S. Army infantry soldier domains. The discussion goes beyond the traditional topics of graphics and game engine technology and delves deeper into concepts of the importance of narrative and the role of affective computing for buy-in by the trainees.

CHAPTER TWO: LITERATURE REVIEW

Chapter Two Summary

Virtual Environment training systems are not a "one size fits all" solution to military training. Spain, et. al., even go so far as to describe this approach to individual and team training as "inefficient" (Spain, Priest, & Murphy, 2012). This is further supported by the individual pockets of development within the U.S. military to produce specialized virtual training environments to satisfy the needs of a community of interest. The potential for virtual environments to enhance military training is widely recognized, however the guidance for the employment of this technology is almost non-existent. The technology must be applied carefully and intelligently so that the trainee may gain the most utility from its use (Reynolds, 2009).

Policy

There is a conflict in the military between the training community's need to provide correct training and the acquisition community's requirement to procure a virtual environment training system in a cost effective manner. This conflict is exacerbated by the large variety of available virtual environment trainers from industry and an institutional culture of a "one size fits all" attitude towards training and education by the higher level decision makers in the military. A key question the military training community needs to answer is which training activities are appropriate for which virtual training solutions (Stanney et al., 2013).

The U.S. armed forces are composed of tens of thousands of war fighters with a diverse set of military occupational specialties (MOS) that vary depending on the branch of the military they are affiliated. Although there is an expectation of a high standard of performance, there are different task proficiencies, different leadership, different operational experiences, and different maintenance of skills. This furthers the argument that a "one-size-fits-all" approach to military training is less effective (Spain et al., 2012).

As a training community, the U.S. military is quite conservative and resistant to change. This is understandable as many of the training activities have mortal consequences. There are certain skills the soldiers absolutely must train for as lives depend on their performance. However, there are financial tradeoffs that must be made when training such a large population. In the current fiscal climate of sequestration, military budgets are expected to constrict over the next decade and cost cutting measures are going into effect across the board. The U.S. Army Training and Doctrine Command (TRADOC) has produced a document called the U.S. Army Learning Concept for 2015. Inside this document, a clear need for distributed and online training capability needs to be developed for the soldier. This document calls for adaptive and flexible training in a virtual environment that can exercise critical thinking skills across a wide range of training activities (Morton, Lucious (Department of the Army, HQ Deputy Chief of Staff, 2011)).

Virtual Environments for Education and Training

Games and simulations share many attributes with some games even being created from simulation engines. Training simulations attempt to recreate some kind of representation of an operating environment (Cannon-Bowers & Bowers, 2009). Simulations do not share some gaming features such as a fantasy component (Charsky, 2010). Simulations that use a game engine for professional training and education use are called “serious games.” The serious game simulation is able to provide an experiential platform where users are able assume a role that requires professional decision making that mimics situations in the real world. The value of the serious game is derived from how closely the simulation is able to provide a realistic experience (Sterling, 2003). Most serious games are used to provoke higher level cognitive abilities and promote critical thinking skills through the presentation of open ended scenarios where the trainee must apply previously learned skills.

The entertainment industry and the training community have many of the same interests. In both cases, believable and realistic scenarios are desired to suspend belief and create immersive environments. The video game makers wish to use this immersion to promote further play and foster customer satisfaction. The training community's desire is to leverage this immersion to promote retention and learning.

The attitudes toward pursuing these interests are somewhat different. The gaming community values frame rates and user interfaces while the training community is concerned with correct physics and accurate behavior models. The gaming community will take shortcuts with physics. For example instead of using ballistics models for weaponry the games will use simple hit-point subtraction calculations for enemy damage. In the same scenario, a modeling and simulation approach is used by the training community to ensure a proper probability of kill is calculated using validated and accredited mathematical models, otherwise known as a constructive simulation (Mcalinden, Clevenger, & Rey, 1998).

Until recently, the blending of these two approaches was computationally prohibitive. High accuracy simulations were often two dimensional presentations and had no immersion, not useful for activities such as mission rehearsals. With today's more aggressive personal computing hardware available at low cost, it is possible to exploit the desirable attributes of both technologies.

Since 2000, many attempts to connect commercial video gaming engines with constructive simulations have produced mixed results. Early attempts found computational limitations in the entities that could be accurately represented, either in their numbers or their complexities.

Additionally, game engines lack robust external communications mechanisms to accept connections from standard modeling and simulation interfaces such as the High Level Architecture (HLA) or the Distributed Interactive Simulation (DIS) specifications.

More recently, successful game engine and constructive simulation integration efforts have been demonstrated. The Unreal Engine (EPIC Games, 2014b) has been successfully stimulated by entities via HLA in a simulation that required external artificial intelligences to perform decision making in a border security demonstration (Richards & Porte, 2009). Additionally, the Tactical Language and Culture Training System created by the University of Southern California and funded by DARPA is used by the U.S. Marines to teach Iraqi and Pashto culture and language skills (Johnson, 2007).

The commercially available off the shelf (COTS) game engines offer many benefits to the military. They present a launching point for a simulation, meaning the software development costs to develop the core capability have already been paid for by industry. The military can create customized game content for the engine at relatively low cost and risk. Even if the military has to license the game engine, this is still more cost effective than developing a new engine from scratch.

The COTS game engines often encourage the user community to create customized content. This extends the life of the game platform and encourages use. The military can leverage this customization capability to create operationally relevant scenarios and content. The tools are often sophisticated and allow for terrain, culture data, building, opposition force composition, and after action review. This capability is known in the gaming community as "modding" or modifying the game. The military can leverage this modding capability as use it as a mission editor to quickly and easily produce training material in a timely manner, responding to evolving situations (Fong, 2006).

There are also challenges to using COTS video games for military training. The realism of the games may not be adequate to properly represent the operational environment. For example, Beal's 2009 study at the Ft. Benning Maneuver Captains Career Course states that the soldier's perceptions of the DARWARS *Ambush* were poor and had little training value due to the lack of ability to react realistically to enemy contact (Beal, 2009). The major lessons learned in this study were that the simulators need to be flexible enough to allow trainees to exercise free will in decision making capabilities. If the simulator does not offer the trainee the options or abilities to react to a situation, the trainee will exhibit a poor opinion towards the training.

Since the games are focused on entertainment, they may take shortcuts when implementing a physics routine, for example. The game companies are very reluctant to license or provide the military access to the game source code, so the likelihood of the military being able to modify that code with AMSAA (Army Materiel Systems Analysis Activity) accredited physics models (called Physical Knowledge Acquisition Data - PKAD) is very small or very expensive. Some deviation from perfect realism is understandable and expected, but is highly dependent on the training activity and situational use.

Another challenge is the background of the trainees. Some of them may be maverick gamers and can pick up the systems with little or no training. Some of them may be completely naive' and will require significant system training before even attempting operational training. This variability is a risk and may be disruptive (Fong, 2006).

The reason computer games have become so successful in the past 20 years is they have the ability to captivate and engage users. Video game manufacturers have produced titles specifically designed to draw players into the game with increasing sophistication (Dickey, 2005). The use of games for educational purposes is not a new concept. Leveraging the well-known engagement properties of the games, educators have tried numerous tactics to use games in a training forum. Video games use narratives, or storytelling, as a way to engage and maintain interest.

Games can be used effectively to augment skills and successfully convey complex concepts (Wray, Laird, & Nuxoll, 2005). Games can be used to affect the attitudes and motivations of a person as well as adjust perceptions. Games can be used to change behaviors and convey motor skills. Of particular interest to the U.S. Army is the ability of games to convey soft skills and social awareness training (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). When compared to the more traditional tools and methods of training, namely slide presentations in a classroom, video games can be more engaging and enhance the motivation of the students to learn (Brusso & Orvis, 2013).

Although military simulations are not fantasy based, the inclusion of a compelling and convincing narrative will enhance immersion and trainee understanding. Conle outlines a number of outcomes associated with the use of narratives in an educational setting. The narrative will provide a frame of reference for the training and set up the scenario for deeper comprehension. Multiple interpretations of the story will provoke discussion and lead to an increase in interpretive competence. Since the narrative is part of an experience, recall is enhanced. The students are active in their participation (Conle, 2003). All of these outcomes are desirable in infantry soldier training.

Simulation based training has been used to present personnel with scenarios that include dangers a live training event would be too dangerous to perform or too costly (or both). Unfortunately the high fidelity training systems are also expensive to create and maintain, as well as too costly to be used on a large scale and are usually reserved for special occasions, rare training events. Bowers et al and Morris both discuss the need to better leverage the lower cost desktop simulators to extend the reach of simulation based training into the military training curriculum (Bowers et al., 2013).

The presentation of the narrative has been shown to be an important aspect of immersive simulation based training. Morris' hypothesis was a trainee could be "pre-primed" to better accept lower fidelity training through the use of a cinematic narrative. Morris was able to show that the cinematic introductory narrative presented to a trainee was able to dramatically increase the trainee's stress and also show better performance within the simulation (Morris, Hancock, & Shirkey, 2004).

Bowers et al. performed an experiment designed to test the hypothesis that a text only presentation of a narrative before a game based training session would be equally as affective as a more expensive cinematic narrative. Their group attempted to replicate a previous experiment performed by Morris in 2004, which showed a cinematic narrative affected training outcome. Unfortunately this experiment was not able to show an increase in stress or performance. The conclusion is a text based narrative presented before the training had no effect on the training outcome. Although much cheaper than the high fidelity alternatives, the costs associated with producing a cinematic narrative are not trivial. Bowers et al. maintain the hypothesis is still worth pursuing with other multimodal cues (Bowers et al., 2013).

Video games for training have been studied for many years and five criteria have been researched by Orvis, Horn, and Belanich which are relevant to assessment of trainee performance and training effectiveness. The first criteria is how goals are managed by the trainee. The goals must be planned and monitored during the training, with adjustments made along the way. This is known as metacognition and it allows the trainee to actively keep up with their progress. The second criteria is the amount of time in the system (Orvis, Horn, & Belanich, 2008). Brown was able to show a positive correlation between the amount of time a trainee spent in a system and the amount of knowledge acquired (Brown, 2001).

The emotional opinion of the trainee must be accounted for. The amount of satisfaction a trainee derives from the training is the third assessment criterion. The fourth criteria is the usability of the training system. The entertainment industry has done a service for the military in vetting and discarding poor designs and offering the commercially successful interfaces for military use. Put simply, if the training system is difficult to use then the trainees may experience reduced motivation to complete or do well.

The fifth criteria for video game training assessment and effectiveness is the performance of the user. Testing to determine how much knowledge is acquired during the training as well as after the training can indicate the effectiveness of the training system (Kozlowski et al., 2001).

Effectiveness of training is always a concern as well as the application of which training activities are appropriate for electronic gaming. Further, the kind of electronic game is also in question. Sitzman was able to show positive instructional effectiveness in a large sample of over 6000 participants using a computer based game with improvements in procedural knowledge and retention (Sitzmann, 2011).

A study conducted by Pennell in 2003 with the British Army attempted to investigate the effectiveness of the Half-Life game engine for the training of building clearance procedures. The conclusions showed the game based training in a positive light. The findings suggested the personnel trained via the game based trainer had better decision making abilities, cleared the rooms faster, and were more judicious in the use of grenades. Unfortunately the study was weak due to the low number of participants and very small sample size (two) of the teams. Additionally all data collected was subjective (Pennell, 2003).

There are many factors that may affect the effectiveness of training with electronic or online games. Orvis found that prior gaming experience, particularly prior experience in the genre of game, was a significant positive factor in the training performance (Orvis, Horn, & Belanich, 2009). Orvis also found that unrealistic expectations for the end goal or outcome of the game had a negative impact on the training performance (Orvis et al., 2008). This provides an interesting insight into proper development and deployment of a game based training system.

Infantry Soldier Training Effectiveness Utilizing Virtual Environments

Ultimately the goal of using virtual environments for infantry soldier training is to encourage skill transfer and promote retention in a cost effective manner. Traditional training methods incur significant costs in the logistics and travel of the personnel involved. Training infantry soldiers in virtual environments offer solutions to some of the challenges posed by live training.

Transfer

There is a significant amount of literature available on the subject of transfer. It is widely described as the application of the products of training, such as knowledge and skills, to the operational environment in which they would be normally used (Alexander, Brunyé, Sidman, & Weil, 2005). Additionally, knowledge transfer in the context of an organization has been defined as a process through which one unit is affected by the experience of another unit (Argote & Ingram, 2000). The common theme that most transfer definitions share is that acquired knowledge is successfully applied from one domain to another. Every training activity, no matter the forum or modality, has the goal of transferring lessons in the synthetic "classroom" environment to the non-deterministic "real-world" environment.

For example, consider the movie "Karate Kid" from 1984. The character of Daniel (played by Ralph Macchio) was given the task to wax his instructor's car. His instructor, Mr. Miyagi (played by Pat Morita), gave Daniel specific instructions for the procedure to wax the car which entailed circular motions of the hands. Daniel was later instructed by Mr. Miyagi that this motion was actually training for a martial arts movement that would block an attacker's punch. Once Daniel made this connection, a transfer of knowledge from the theoretical to the practical, he was able to later defend himself from antagonists later in the movie (Avildsen, 1984).

Training activities may be considered positive, negative or neutral (Alexander et al., 2005). This is determined based on the performance of the trainee in the operational environment, meaning positive training improves performance while negative training has a negative effect on performance.

In its report to the United States Congress, the Government Accountability Office criticized the Army for not using standardized methodologies to measure transfer in simulation based training activities (Pickup, 2013). Further, the methodologies that are employed often do not follow a quantitative approach, rather simply the subjective opinions of the local commanders or observer controllers (Insinna, 2013). It is important to note this is a difficult activity and the effectiveness data collected is only valid for the type of training conducted, the task or tasks trained, the measurement methods, and the measurement scales.

Transfer of training is a function of percentage of transfer where the degree to which learning has taken place is measured in some way. The learning of a particular task is achieved by prior instruction, practice, or study and then the learner demonstrates the task and is measured by expert on ability (Roscoe & Williges, 1980). Equation 1 shows the general formula for percentage of transfer:

$$Percentage\ of\ transfer = \frac{L_X - L_0}{T - L_0} \times 100 \qquad (1)$$

L_0 is the average learning on a naïve control group, where this group has had no prior training on the task. L_x is the average learning of the control group with prior training on the task. T represents a perfect score. Roscoe notes it may not be possible to precisely determine T.

A second formula for percentage of transfer is much more interesting as it makes provisions for a symmetrical transfer curve with both positive and negative outcomes. It is possible to calculate negative training outcomes with Equation 2.

$$Percentage\ of\ transfer = \frac{Y_0 - Y_x}{Y_0 + L_x} \times 100 \qquad (2)$$

Y_0 and Y_x are the same as L_0 and L_x in Equation 1. The major issue pointed out by Roscoe with these equations are they do not take into account just how much practice on the task is performed which may skew results.

Roscoe goes on to discuss the concept of incremental transfer, or the effectiveness prior training on a task has on the subsequent training effectiveness. In other words, how practice affects skill acquisition. An acceptable method for showing the effectiveness of pre-training is the cumulative training effectiveness function (CTEF), Equation 3.

$$\text{CTEF} = \frac{Y_0 - Y_x}{X} \tag{3}$$

Y_0 in equation 3 corresponds to Y_0 in equation 2 where the variable may be errors, trials, or times to complete a task by the naïve control group. Y_x is the measure for the control group with practice or some other prior training. X corresponds to the known error, trials, or times to complete a task by a separate experimental control group with prior practice. X could be considered an established baseline (Roscoe & Williges, 1980).

The U.S. military prefers a cumulative transfer effectiveness function to attempt to take into account learner prior exposure and practice (Fletcher, 2009). Fletcher also describes the U.S. military's desire to implement simulation based training systems that can train tasks that escalate in complexity. This is key to training to an operational environment that is also complex. Simulation is seen as a possible mechanism to shorten years of traditional training required for mastery.

In order to satisfy the desire to establish time savings through simulation, the return on investment must be established. Additionally, any diminished returns must be established to prevent over-investment in a simulation technology. Using the techniques described in this section, estimates may be established for the application of knowledge acquired in a simulation based trainer to the operational environment.

Presence and Immersion

Singer and Witmer describe *presence* as a personal feeling, a subjective experience, in which a user of a virtual environment believes they are part of that environment. This feeling or belief is a physical detachment. They define *immersion* as the mental state of perception in which a user of a virtual environment is interacting with that environment. Singer and Witmer then go on to describe a strategy for the practical measurement of immersion through their Immersive Tendencies Questionnaire (ITQ) and a Presence Questionnaire (PQ) (Witmer & Singer, 1998). According to Singer, the idea behind these attempts to measure immersion and presence was to answer the question "What does it mean to be immersed?"

Slater attempts to take a more objective look at presence and immersion. Slater's notion of presence employs a "sense of being there" in the virtual environment, the extent to which the virtual environment dominates the user's real experience, and the extent to which a user remembers the virtual environment as a place rather than a presentation (Slater, 1999). All of these aspects of presence have been measured subjectively through experimental studies involving observation and user interviews. Slater understands the delicate balance between measuring immersion and destruction of the immersed experience through the act of polling the participant with questions such as "How immersed are you right now?"

The controversy surrounding Singer's approaches to measuring immersion and presence versus Slater's approaches highlight the differences in subjective versus objective analysis of the data collected. Additionally, the user-centric focus of Singer's work is a contrast to the system-centric focus of Slater's work. Whether intentionally or unintentionally, users are not always truthful or accurate in their self-assessment. For example, in Singer's PQ a number of users are asked "How compelling was your sense of moving around inside the virtual environment?" The acceptable responses are provided in a scale ranging from "Not Compelling" to "Moderately Compelling" to "Very Compelling". Given the extreme hypothetical case where all participants respond "Very Compelling" then the virtual environment is statistically deemed 100% Very Compelling. The issue Slater takes with this approach is that "compelling" is not adequately defined to provide a useful metric for the virtual environment.

Slater therefore comes to the conclusion that immersion and presence require a much more technical definition such that variables may be assigned and manipulated. For example, a user's sense of reality is described by four factors. Inclusive (I) is the extent to which a user's real world is excluded. Extensive (E) is the amount of accommodation of the senses. Surrounding (S) is a measure of the user's view frustum. Lastly, vivid (V) is a record of the characteristics of the technology used to provide the virtual environment (Schubert & Friedmann, 2001).

Singer focuses on the user and attempts to identify their previous experiences or biases. The ITQ is intended to try to identify how susceptible to immersion a person may be. This susceptibility may be key to acceptance of the virtual environment and therefore a different level of immersion than someone who is resistant or hostile to the experience. These levels of immersion and presence are referred to as "degrees" in the literature and Singer is interested in user focus and amount of attention a user is devoting to the virtual environment.

Slater attempts to spread the responsibility of immersion broadly. Not only is susceptibility of immersion important, but also the technical presentation of the virtual environment. Slater's work sets out to define as many dimensions and combinations as possible any affectations to a user's virtual environment experience from a technical standpoint. For example, Slater insists that with all other things being equal, a virtual environment that provides the user with a wider field of view than another virtual environment, then the first *must* be more immersive (Slater, 1999).

Stress

Research on real tasking such as vehicle operation, industrial tasks, and military operations has shown performance is frequently stressful (Matthews, Szalma, Rose, Neubauer, & Warm, 2013). The tasking may be perceived as stressful due to factors such as workload, time to task completion, or even probability of failure. Matthews, et. al. point out that even environmental factors such as noise, hot, cold, or limited endurance can be involved. There are social factors, such as team cohesion that can add to an individual's stress. The Dundee Stress State Questionnaire (DSSQ) was developed to assess the various subjective states associated with performance and stress, specifically the factors associated with task engagement, distress, and worry (Matthews et al., 2013). The DSSQ is a validated measurement tool with primary scales being mood, motivation, and cognition (Matthews et al., 1999, 2013).

Task engagement is defined by three state factors: energy, motivation, and concentration. Enthusiasm and interest in the task are contrasted with fatigue and apathy. This factor attempts to gauge the user's motivation. Mood is characterized by three Boolean discriminators, energetic arousal, tense arousal, and hedonic tone. Energetic arousal refers to the person's fatigue such as being vigorous or tired. Tense arousal refers to their state of being nervous or relaxed. Hedonic tone refers to their agreeability or being in a pleasant versus unpleasant mood. Motivation as assessed by the DSSQ and relates to the interest the participant has in the training or task at hand.

Cognition is recognized as being the most difficult state to assess due to the many factors relevant to its construct. According to Matthews, et. al., specific beliefs and attitudes are not applicable and the dimensions are represented by general states of being. For example, cognitive states are affected by awareness and are persistent over time. The worry state is characterized by numerous intrusive thoughts, often negative. The worry state can be measured by the number of distractive thoughts, not the actual thoughts themselves.

Current Use of Virtual Environments for Infantry training

The U.S. Army already uses virtual environments for small unit training activities and has produced at least three classes of video game technologies for infantry soldier training. These classes include the first person shooter, the massively multiplayer online game, and the virtual world. Although the technology may overlap, each of the types of video game have distinct attributes.

The U.S. Army has limited its virtual training activities for infantry soldiers to small unit dismounted operations, combat mission rehearsal, and live training mission rehearsal. (Kaber et al., 2013) Large unit or collaborative small unit training is not performed. This is primarily due to the kind of virtual environment technology it has aligned itself with, namely the first person shooter style game engine. The first person shooter is a type of video game where a three-dimensional world is presented to the user. The user is able to navigate the world with the view of the computer screen as the view through the eyes of their character avatar (Schneider, 2004). The users is outfitted with various weaponry and gear and the goal of shooting opponents while moving from one part of the game area to another.

Live training for infantry soldiers is commonly performed via MOUT (military operations in urban terrain) presentations. The two main elements of MOUT training is to learn how to neutralize enemy threats and how to avoid taking unit casualties (Hale, Stanney, & Malone, 2009). There are many MOUT facilities located around the world which typically a composed of a classroom component and full scale mockup of a representative real world urban environment. The mockup can be as simple as a single room or as elaborate as an entire town. The typical training cycle for infantry soldiers is to have tactics and strategy taught in a classroom setting (usually with a slide presentation) and then the soldiers are sent to a mockup site for rehearsal style training. The training is observed and rated by commanding officers.

The U.S. Army is interested in reducing the costs and increasing the effectiveness of the live training component of a soldier's training cycles. Live training is expensive and time consuming and virtual environment based training is seen as a possible way to augment live training to reduce costs. Early efforts at virtual environment team training focused on simple tasks such as single room clearing. In 2001, Lampton and Parsons described a research system called Fully Immersive Team Training (FITT). The FITT was created and designed to support examination of virtual environments for team training effectiveness. The FITT was an ambitious system composed of an individual combatant simulator, a synthetic combatant entity server, the mission control station, an independent audio system for radio communications simulation, and a data collection / after action review system. The very first uses of the FITT were to study strategy for a prescribed scenario by studying training manuals and procedures, then practice the missions in a virtual environment (Lampton & Parsons, 2001). The performance measures of this early study included timing and accuracy of the tasks assigned to the trainee.

An important example of a success of the use of virtual environments for training is in the conveyance of spatial knowledge. Spatial knowledge is a key component to many military domains, including piloting vehicles (air, land and sea) and dismounted infantry land navigation. Spatial knowledge also has a limited shelf life and virtual environments provide a cost effective means of providing training maintenance (Stanney et al., 2013).

To accomplish infantry soldier skills training, a serious game must satisfy a number of basic criteria: (1) the virtual environment of the game is a mirror world that simulates reality and the real world, but actions have no "real" consequences. This is a vital attribute of gaming for training that allows for safe and cost effective action. (2) The game must consistently follow a set of virtual world rules. If the training is to be correct, the virtual environment must replicate what is being trained faithfully. (3) The game must have a goal or exit criteria. The player needs to know when they are done and if they are successful or not, meaning there is at least one correct solution (Burgos, Tattersall, & Koper, 2007).

Commercially available off the shelf (COTS) games have been applied to military training since the 1990's. Until recently realism was a barrier for COTS gaming in collaborative and team training applications. Older games were fine for the entertainment industry, but didn't have enough realism to be used for infantry soldier training (Petroski, 1985). Today, games such as "Call of Duty" offer a highly realistic presentation for infantry soldier training.

Massively Multiplayer Online Gaming

Massively multiplayer online games (MMOGs) for educational use have been studied for many years. MMOGs are virtual entertainment environments with a two or three dimensional presentation. Users are encouraged to customize an avatar and role play through the game story.

MMOGs are virtual environments endowed with the ability to motivate players to interactively achieve goals, actively engage in problem solving activities where they attain greater and greater expertise to progress through the game. They often provide a means to follow a story line and cooperate with other players to form groups that multiply their force or strength to progress even further than they would as individuals (Voulgari, Komis, & Sampson, 2013).

MMOGs are often endowed with an economy where users can trade or buy items that help them achieve goals. Childress and Braswell integrated a MMOG into a graduate online course.

They were able to show that the environment fostered cooperative learning through the highly social attributes of the MMOG. They were able to set up and observe successful partner and group activities inside the virtual environment. They concluded the environment was able to provide a realistic proxy for a real life meeting experience, leading to positive collaborative critical thinking and problem solving (Childress & Braswell, 2006).

The ties to education and gaming, specifically the MMOG can be made by first examining theoretical framework for the three aspects of learning: cognitive, social, and emotional (affective) as explained by Voulgari, et al. The authors argue for links between traditional learning theories and patterns of activity and within the context of successfully operating within a MMOG.

The cognitive aspect of learning in this domain refers to conceptual and factual knowledge acquisition. This learning can be navigational knowledge about the three dimensional environment, acquisition of skills to operate within the game framework, personal optimization of communications mechanisms utilizing the provided game tools, formalizing strategies to complete missions and attain goals.

The social aspects of learning deal with how to cooperate with other players and operate as a team. Leadership skills are very important with players learning how to coordinate resources, recognize strengths in their subordinates and utilize them to best of their abilities, and mission planning. The MMOGs typically have robust communications mechanisms to promote team play. These mechanisms may manifest themselves in both text based communications as well as voice.

The least understood and investigated aspect of learning in the MMOG is the affective or emotional subject. The players must remain motivated to keep playing. The games must maintain a level of difficulty to maintain interest, but not be so difficult as to be de-motivational. Players will remain vested in the game if they are satisfied with their status, happy with their progress and status, and maintain a positive attitude (Voulgari et al., 2013).

Voulgari discusses a number of learning outcomes related to players in MMOG environments. These learning outcomes include the acquisition of cognitive skills with skill-based outcomes, and emotional (affective) impacts.

A signature feature of the MMOG environment is the integration of quests and missions (problems) for which players must actively pursue solutions. The players must learn the unique social structure of the game, how the objects within the game work and affect outcomes, the overall rules, and general functionality. The games may be quite complex and require intense concentration during times of action. For the players to attain proficiency in the game, they must practice and spend time in the game.

The game designers are often clever and design ever increasing difficulty of the scenarios to encourage the players to continue. The MMOG often present goals to attain or problems to solve that may have many avenues to approach the solutions. These goals and problems are complex and the player must use their knowledge of the game framework to pursue solutions. These open ended problems promote critical thinking skills and encourage the player to continue and set up the proper conditions for learning (Voulgari et al., 2013).

Metacognition is a reoccurring theme in the literature. Metacognition in this context is described as a certain self-awareness possessed by the player. This self-awareness is the knowledge gained by a player to know when to apply a strategy for problem solving. It is a process that involves careful examination of a situation and the selection of a method to attempt to arrive at a positive outcome (Shetty, 2010). Voulgari maintains metacognition is required for the selection of the relevant skills necessary to complete a task within the MMOG. Metacognition is also required for the player or players to try different strategies, promotes connected discussions and conflict resolution in the group. Voulgari further discusses the transition of the skills learned in-game to other games as well as, and arguably more importantly, the real world.

Ang and Zaphiris describe observed social interactions within the MMOG "World of Warcraft". These observations led them to explain the possibility of social learning through eight aspects. These were community mediated learning, in-game social norms, collective knowledge construction, social learning through group reflection, conflict, player goal construction, shared tools for information exchange, homogeneous and heterogeneous social learning (Ang & Zaphiris, 2008).

Community mediated learning has been observed as one of the common forms of social learning. Players (new and existing) will observe the actions of others and pay attention to the outcomes. Emulation of the successful strategies leads to learning new ways to play the game. Additionally, more experienced players may tutor the new players and provide information during an activity. Lastly, the after action reporting of the collective group on what they did and how the strategies compared are a valuable way for the new players to gain competency.

MMOG environments are social in nature and as a consequence a social structure and norms emerge. Players must communicate and coordinate actions in order to successfully progress through the game. The interactions force the emergence of leaders and followers through natural self-organizing groupings. Through coordination and planning, a division of labor and allocation of resources takes place. These groups begin to establish a subset of gaming social norms and form group specific norms to match the hierarchy of the group (sometimes known as a clan).

Key attributes of the MMOG that make it ideal for investigation for training uses are the interfaces which foster social interaction and player communication. Often these games require teamwork to accomplish tasks to advance in the scenario or gain rewards for the player's role avatar. In the early days of MMO gaming, text chat for player to player communication as well as group chatting was essential for team coordination. In military terms, this would be part of a command and control system. Today, many modern MMO games also incorporate voice chatting capabilities to enhance the experience and expedite coordination.

Another key attribute of the MMOG is the three dimensional operating environment with almost photo-realistic graphics. This visual presentation with accurate physics response (gravity, avatar movement, collision detection) provides an easy way for the player to become immersed in the game.

Research into the use of MMO gaming technology for education and training has been performed on a wide variety of domains. For example, Peterson makes a compelling case for the use of MMOG in second language learning. Peterson states that two types of interaction facilitates second language acquisition. The first is interaction is how people work together while dealing with communication problems and the second interaction is a focused attention of a person on the clarification of their communication output (Peterson, 2010). MMOG technology was shown to be a good fit for this type of training due to the inherent social tools provided by the environment.

Another promising use of MMO gaming technology is for leadership skills training. Since the MMO game environment is collaborative and social, it is natural that groups of people will follow an ordered hierarchy of some kind in order to plan and execute events to accomplish some sort of goal. Often in the entertainment setting, these groups are spontaneous and the hierarchy determined by the group. In the military, however, the chain of command is well understood and followed. Games such as StrikeCOM was a MMO designed to study group interactions for military use (de Freitas & Griffiths, 2007).

Virtual World Technology

The virtual world technology represents a subset of the Massively Multiplayer Online Game. The virtual world is usually a persistent world where time proceeds whether the user is logged on or not. In a virtual world, every object is an agent. This means a user has the ability to interact with every object in the virtual world, just like the real world. This subtle, but critical distinction has significant implications for how the virtual world can be used as opposed to the more traditional MMOG.

In a first person shooter (FPS) style game engine or a traditional MMOG engine, the three dimensional models and artwork that is used to populate those environments have a very different ingestion pipeline than a virtual world. In the FPS or MMOG, digital artists and a game designer's must plan well in advance all the mechanics of the game play as well as all of the content composing the game scenario. This can be compared to the production of a cinema movie, with storyboards and a complete list of all models and artwork to go into the game before it is ever built. The levels are populated using a separate "world editor" and then imported into the engine for the players to enjoy. Any changes to the world must be done outside the engine and re-imported. All interactions must be planned for in advance. For procedural based training applications, this is fine. However critical thinking skills requiring non-determinant behaviors and player un-scripted decision making is difficult in this framework.

Uses for this kind of technology have been demonstrated in collaborative engineering and mechanical design, building architecture, data visualization, and more. The ability to interact with any object in a virtual environment without previous planning is thought to be a key ability to provide an unscripted and non-determinant environment giving a user the freedom and ability to make decisions and exercise critical thinking skills in problem solving. This is especially helpful in the soft skills training areas being considered by the U.S. Army for human network analysis and cultural awareness.

The makers of virtual worlds attempt to faithfully represent the experience of a real world as much as possible. The virtual worlds have "world rules", such as physics for collision detection and gravity. The virtual worlds strive for consistency of these rules. For example, a user should not be able to walk through a wall or fall through the ground. The virtual worlds are populated with "avatars" who are people logged in to the system. Some virtual worlds allow the users to customize their avatars to a high degree and this is a popular activity which allows for self-expression (Dev, Youngblood, Heinrichs, & Kusumoto, 2007).

The education community has been aggressively exploring the uses of virtual worlds. Educators are interested in knowing if the virtual worlds can promote active learning better than traditional classroom methods of teaching. Active learning is defined as the technique that requires a student to process and apply concepts and information presented, as opposed to passive learning which would be simply listening to a lecture or watching a movie (Wang & Braman, 2008).

Virtual Worlds Team Training

In the early 2000's, the corporate world had taken serious notice of the virtual world technology and experimented with it as a cost effective environment for distributed and collaborative learning (Nebolsky, Yee, Petrushin, & Gershman, 2003). A study performed by Heinrichs, et. al. in 2008 showed that even using a low fidelity world simulator, team training in the medical domain can be performed in a cost effective manner (LeRoy Heinrichs, Youngblood, Harter, & Dev, 2008). Later, corporate training systems have been developed that leverage virtual world technology to enable leadership training through role play (Kark, 2011).

Modern Virtual Environments and Military Team Training

Even though it has been shown that virtual environments are useful for many forms of military training, the technology has been difficult to deploy widely. There are many barriers to entry for the employment of this technology. Possibly the biggest barrier to adoption of modern virtual environments for simulation based military teat training is information assurance. Military information assurance policy is rigid, well defined, and not well suited for advanced distributed game architectures. Since the main characteristic of a multiplayer online game is the number of users in the game simultaneously, network usage can be randomly heavy. Additionally, to spread the load of the network to accommodate so many users, many network ports must be utilized (Bezerra, Comba, & Geyer, 2012). Military network security rules prohibit large numbers of ports from being open as well as any so called "non-standard" ports. While it is possible to get exceptions for certain specific uses of the virtual environments, these exceptions are limited in scope and don't properly exercise the technology.

Summary of Literature Review

The literature indicates more work needs to be done in the areas of knowledge transfer and which kinds of training activities are appropriate for virtual environments. Research from Ang, Voulgari, and others show there is promise for the transfer of knowledge within specific domains as well as the ability to leverage the social aspects of the virtual environments for learning.

According to Dickey, the engaging aspects of video games should be examined as strategies of instructional design. As the entertainment industry discovers through a process of evolution more and better ways to keep players in the platforms, so should the training and education community monitor their progress. Of particular interest is the idea of including narratives in the instructional design process utilizing computer games. The recipients of the training may find it easier to become emotionally involved with the training if they can identify with the training through the use of a real world example or scenario. This affective approach has been discussed as a possibility to promote engagement and therefore promote knowledge retention and accelerate learning.

A key aspect of this subject underrepresented in the literature search is guidance for when video games should be used in training and education. Not all subjects or activities are suitable for training with a serious game. Further, the decision to use a serious game to train is not Boolean in nature. More research is needed to define which activities or classes of activities are appropriate for serious gaming and which are not. This research can also help to define if there are situations under which the use of video games for training actually cause more harm than good.

Research performed specifically to determine which U.S. Army infantry soldier training will require a methodical examination of existing training subjects and activities. The U.S. Army

Training and Doctrine Command (TRADOC) is the responsible agent for all accredited infantry soldier training. Unfortunately, this is a tremendous undertaking as the entirety of the TRADOC training material should be consulted.

Initial research will focus on the informed selection of certain common training tasks. U.S. Army units will be identified with significant numbers of available soldiers for experimentation. The experiments will be comparative in nature and evaluate the performance and knowledge transferred during traditional TRADOC curriculum versus non-traditional serious game training delivery.

The results of these experiments will provide information with respect to differences in time, cost, proficiency, and maintenance of knowledge between serious games versus traditional classroom presentation (Dwyer, Griffith, & Maxwell, 2011). The differences will result in a way to calculate returns on investment for each of these parameters. An additional result of these experiments will also yield an experimental design for other training and education researchers to follow when adding more training activities to the overall comparison study.

Lastly, when the appropriateness of various training activities for use in serious games can be established, then accreditation can be addressed. Ultimately, a serious game could be created that would be used as both the training and the accreditation for the training activity. The behavior of the trainee in the game would be observed by training officer and their performance assessed with a pass or failing grade.

Research Gap

This literature review has revealed a lack of knowledge surrounding the efficacy of the practical application of virtual world technology for infantry soldier training, specifically ground combat skills training such as room clearing. Due to the current subjective nature of gauging training effectiveness of virtual environments, it is difficult to calculate a return on investment. Lastly, it is difficult to determine comparisons of knowledge transfer between traditional training means and virtual training activities for ground combat skills.

Whether it is labeled virtual world technology, game based virtual environments, or virtual environments, the technology is becoming more ubiquitous in the lives of infantry soldiers. However, the literature is unclear as to where in the ground combat skills training cycle this technology is applied the most effectively. The literature is terse in the proper tasks the technology should be used for training. Further, the assessment methodology is not standardized and is often performed using purely subjective means.

Three questions are addressed in this work:

1) Is there a difference in the training effectiveness of a virtual training system versus traditional classroom means?
2) Are differences measurable for individual soldier performance?
3) Are performance differences measurable for collective activities?

CHAPTER THREE: METHODOLOGY

Chapter Three Summary

Chapter three describes the experimental approach using separate studies, the participants used in the effectiveness evaluations, hypotheses, objectives, hardware and software used in the experimentation, procedures, and statistical analysis methods utilized for the data collected. Additionally, this chapter describes the specific data collection events from January 2014 and February 2015 using Florida Army National Guard units.

Long Term Research Goals: Crawl, Walk, Run

The experimentation described in this study is organized into two data collection events which represent the "crawl" phase of a greater, long-term, infantry soldier training effectiveness evaluation. The crawl phase is the comparison of a single task trained by conventional classroom means to a simulation based training system in a virtual environment. The comparison is accomplished by taking the two groups of participants who are trained by the two different training conditions and using Observer Controller (OC) Subject Matter Experts (SMEs) to assess their performance. The training condition chosen for this study is a room clearing task that requires a fire team composed of four soldiers to enter a room and search a room. The participants are assessed both for individual performance as well as group performance.

The walk phase will be conducted after the completion of the initial soldier training performance effectiveness evaluations. The walk phase will include collective training tasks that include higher echelons of soldiers, beyond the fire team level tasks in the crawl phase. These activities consist of two parts and will begin in the winter of 2016. The first phase will introduce two training conditions to be completed sequentially by the participants. This task stacking approach begins to approximate real world missions. The second phase will use a single training condition, but require multiple fire teams to work together to accomplish an objective. This cooperative team training environment also begins to approximate real world missions.

The run phase combines the multiple tasks to be accomplished with the cooperative behaviors of multiple teams. An actual set of operational orders is not specific on exactly which tasks the infantry soldiers will be expected to perform. The orders are open ended and the soldiers must rely on their training to handle fluid and ever changing situations. In an actual operating environment, the soldiers must be able to respond in an agile manner.

The completion of the run phase will not provide the U.S. Army with an over-arching guidance for how to employ simulation based training nor will it provide broad estimates of return on investment. This study has a narrow focus on specific set of tasks and any recommendations presented should be taken in context. Rather, the purpose of this study and the overall long term effectiveness evaluation is to provide guidance for how to perform the studies needed to collect the broad base of data for use in higher level acquisition and policy decision making.

Crawl: Two Experiments

The data collection events represent the presentation of a single ground training task to two groups of soldiers with differing expertise. Group #1 represented a blended group of experienced soldiers and Group #2 will represent a novice pool of soldiers who have never performed the actions trained in the study in combat. The training condition chosen for this study is a room clearing task that requires a fire team composed of four soldiers to enter a room and search a room. The participants are assessed both for individual performance as well as group performance.

Room Clearing

Room clearing represents one of the most common tasks performed by an infantry solder, and the author has been told by many soldiers that this task is also one of the most dangerous. Although this is a collective task, each of the individual positions in the task is assessed independently. This allows for both an individual performance assessment and a collective assessment.

There are four performance steps required to conduct a room clearing task (*FM 3-21.8 The Infantry Rifle Platoon and Squad*, 2007). During step one, the unit leader takes a position in the area that enables them to best control the security and clearing teams. The unit leader directs a team to secure hallways, corridors, and through-fares outside the room with appropriate weaponry. The clearing team leader selects a position from which to best control the team outside the room. The unit leader then gives a signal to clear the room.

Once the unit leader gives the signal to clear the room, step two commences. This is a very rapid action with four specific entry sequences. The first soldier enters the room and immediately eliminates any threat (figure 1). The soldier moves along the path of least resistance, either to the left or right to a point of domination (one of the two corners) and continues into the room.

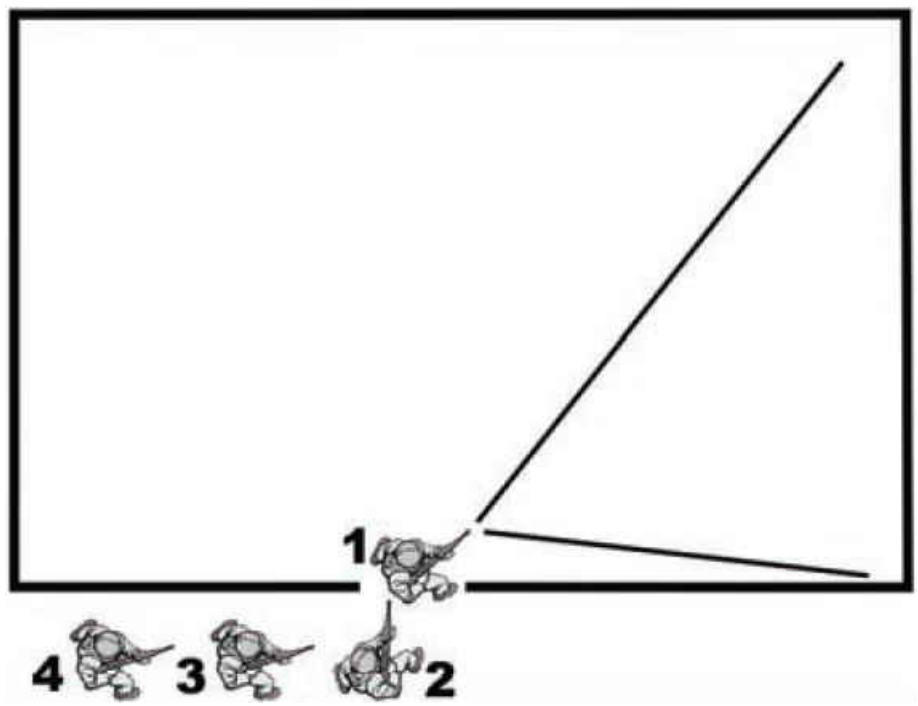

Figure 1. Soldier one entry.
Source: FM 3-21.8 The Infantry Rifle Platoon and Squad
http://armypubs.army.mil/doctrine/DR_pubs/dr_a/pdf/fm3_21x8.pdf

The second soldier enters the room simultaneously with soldier 1 and moves in the opposite direction, following the wall to the opposite point of domination (figure 2). Threats are eliminated in the area while the soldier moves to the opposite corner.

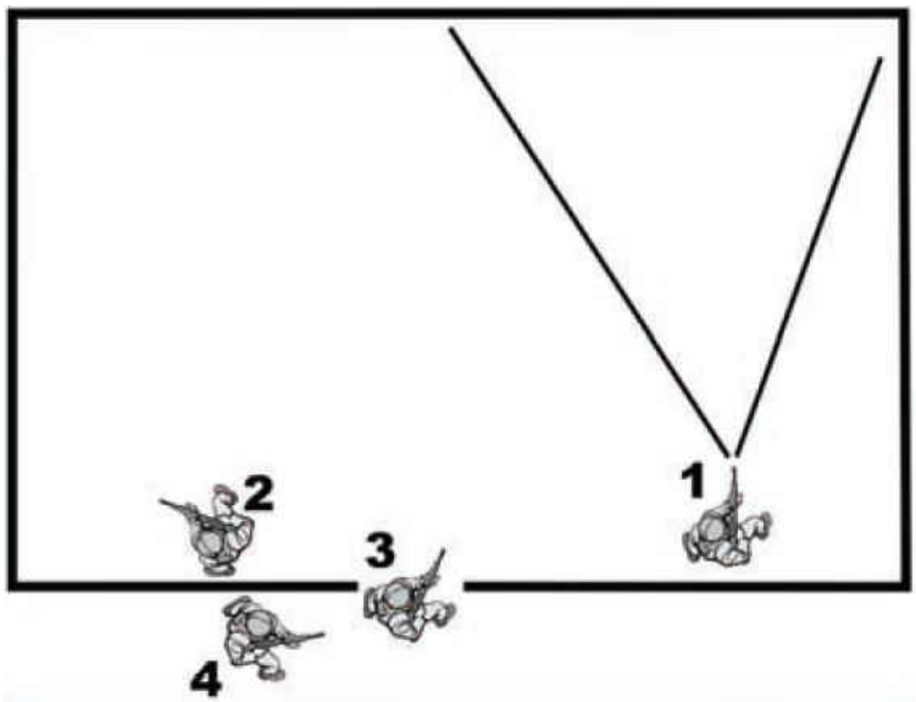

Figure 2. Soldier two entry.
Source: FM 3-21.8 The Infantry Rifle Platoon and Squad
http://armypubs.army.mil/doctrine/DR_pubs/dr_a/pdf/fm3_21x8.pdf

Soldier three moves in the opposite direction as soldier 2 and moves at least one meter into the room. Soldier 3 takes a position that dominates this sector and watches for threats to eliminate. Lastly, Soldier 4 moves in the opposite direction as soldier 3 and ensures the doorway is clear. Soldier 4 takes a position that dominates this sector and watches for threats to eliminate (figure 3).

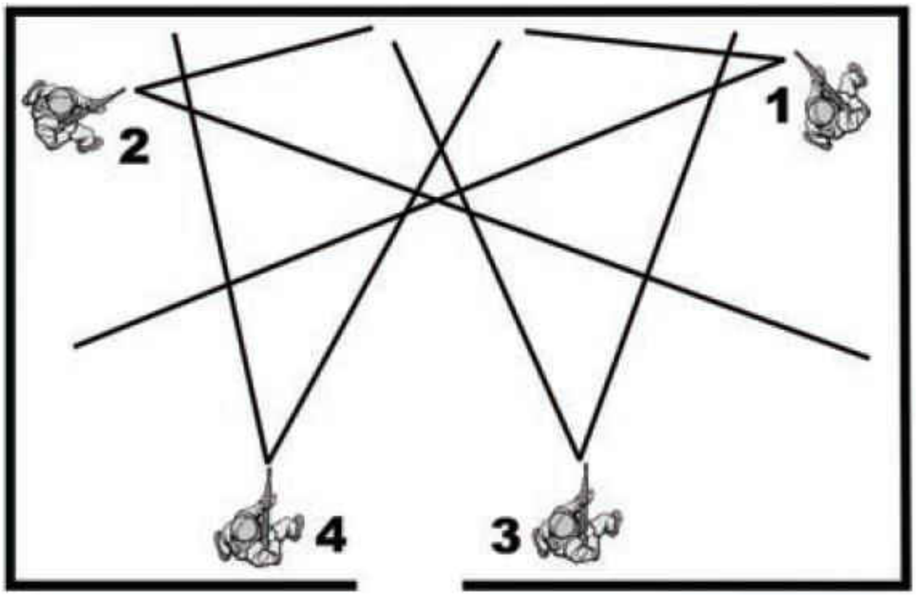

Figure 3. Soldiers 3 and 4 take position.
Source: FM 3-21.8 The Infantry Rifle Platoon and Squad
http://armypubs.army.mil/doctrine/DR_pubs/dr_a/pdf/fm3_21x8.pdf

The SMEs rating the individual and collective performance followed a 4 step rubric. Step one of the rubric (entry phase) was to assess the speed of entry, removal of self from the entry area, follow the path of least resistance and flow of movement. Step two (eliminate threat phase) was to maintain correct sector of fire throughout the flow. Step three (position of dominance) was to assess the soldier's ability to move to the correct position of dominance for their position in the entry team and for the team leader to announce "CLEAR". Step four (Consolidation and Reorganization) is to assess the team's ability to report ammunition, casualty, and equipment status (ACE report).

Experimental Goals and Objectives

The overall goal of this series of studies is to determine if there is any impact on training effectiveness utilizing virtual training methods versus traditional means. The first objective of these studies is to determine any performance outcomes of infantry soldiers employing traditional training versus GBVE training methods for a specific task. The second objective of this study is to determine if team performance outcomes of soldiers are affected when provided GBVE training methods as compared to traditional means. The use of experienced and novice soldiers may also reveal training effectiveness differences in each population, thus the third objective of this study is to explore if whether one population may benefit from GBVE training methods more than the other population. This may also provide indications as to where in the training cycle the GBVE training is most appropriately performed for this task.

To accomplish these goals, this study performs evaluations of the effectiveness of the virtual environment training. A simple task was chosen so that the performance can be performed explicitly. The more complex the task, the more difficult it becomes to define the measurement of the performance in the system (Salas, Rosen, Held, & Weissmuller, 2008). The example task used for this evaluation is a room clearing exercise. The virtual world technology chosen for this experimentation is the Open Simulator due to its flexible licensing, low cost, and the relative ease of scenario development (Maxwell & Ortiz, 2013).

It is anticipated these studies will assist in the development of design methodology to provide recommendations for evaluating candidate GBVE training applications by the military when considering the technology for an augmentation or replacement to existing training means.

Experiment Challenges

In order to obtain domain relevant data, the author made the deliberate decision to only use U.S. military personnel in this study. Access to infantry soldiers was a challenge. The typical source of soldiers would be Ft. Benning, Ga, however there is intense competition for their time and there is a need to ensure they accomplish the training they are there to receive in the limited amount of time allotted to them before deployment. An unorthodox solution was to use the Florida Army National Guard. Since the U.S. military has maintained a high level of warfighter readiness due to persistent conflict over the past decade, the National Guard represented a largely untapped pool of experienced infantry soldiers. A second study was performed using University of Central Florida Reserve Officer Training Corps (ROTC) students and represented a naïve pool of participants with appropriate domain knowledge and had appreciation of the training presented in the experiment.

These studies included a team performance comparison component. Another challenge is the ability to obtain and coordinate enough teams from the required soldier populations to provide statistically significant amounts of data for proper analysis. Table 2 shows the population source by unit, anticipated individual and collective numbers of participants. By using the 2/124th Apache Company and the UCF ROTC detachment, an estimated 128 participants will be available for these studies. This many participants provides a power of over 92% and is discussed in detail in chapter five.

Table 2. Army Unit Source and Numbers of Individuals and Fire Teams

Subject Source	Individual n	Collective n
FLARNG 2/124th	64	16
UCF ROTC	64	16

Study #1: Experienced Population

Study #1 Participants

The Apache Company 2/124th Florida National Guard is located in Leesburg, Florida. This group of soldiers was comprised of 64 male reservists and were used in 16 fire teams composed of four soldiers each. The participants ranged in age from 19 to 35 years with a mean of 25.22 and standard deviation of 3.8. The participants were recruited through their monthly drill exercise and each squad of two fire teams was randomly assigned to the virtual or live condition.

This population is considered "experienced." The population had a mean of 4.49 years of service, with a range of 0.58 to 11 years. This population ranged in number of deployments from none to three, with a mean of 0.69. Lastly, 94% of the population had prior room clearing training.

The soldiers were provided with consent forms and were given the opportunity to review the experiment objectives. Investigators were available to answer any questions. The soldiers were asked to sign the consent forms to indicate they understood their participation was voluntary.

The first study was conducted in January of 2014. The hypotheses tested during this study are discussed below.

Hypothesis 1: Performance Effect of Different Training Conditions of Individual Experienced Soldiers

Hypothesis one explores the primary objective of this study and is posed to determine if there is a difference on the performance of experienced soldiers who are trained using traditional means versus a game based virtual environment. It is hypothesized that "*for the task of clearing a room, there is no difference between the performances of experienced soldiers who have been* *trained using traditional classroom means versus experienced soldiers who have been trained using game based virtual environments*."

The independent variable for this hypothesis is the training condition and the dependent variable is the evaluation provided by the subject matter expert(s). The list of variables for $H1_0$ are shown in Table 3.

Table 3. List of Variables for Hypothesis 1

Variable	Type	Measurement Method	Measurement Scale
Training Condition	Independent	n/a	
Task Performance	Dependent	Subject Matter Expert Evaluation	Pass/Fail

Hypothesis 2: Performance Effect of Different Training Conditions of Teams of Experienced Soldiers

Hypothesis two explores the second objective of this study and is posed to determine if there is a difference on the performance of teams of experienced soldiers who are trained using traditional means versus a virtual environment. It is hypothesized that "*for the task of clearing a room, there is no difference between the performances of teams of experienced soldiers who have been trained using traditional means versus teams of experienced soldiers who have been trained using game based virtual environments.*"

The independent variable for this hypothesis is the training condition and the dependent variables are the team's task performance and the relative rank of the team's performance. The task performance and ranks are provided by subject matter expert(s). The reason for the rankings is to allow for additional non-parametric statistical analysis using Wilcoxon techniques (Mendenhall & Sincich, 2007). The list of variables for $H2_0$ are shown in Table 4.

Table 4. List of Variables for Hypothesis 2

Variable	Type	Measurement Method	Measurement Scale
Training Condition	Independent	n/a	
Team Performance	Dependent	Subject Matter Expert Evaluation	Pass/Fail
Team Rank	Dependent	Subject Matter Expert Evaluation	Integer

Study #2: Novice Population

Study #2 Participants

The University of Central Florida ROTC detachment is located in Orlando, Florida. This group of novice soldiers was comprised of 64 male and female officers in training and used in 16 fire teams composed of four soldiers each.

The participants ranged in age from 18 to 28 years with a mean of 19.9 and standard deviation of 1.84. The participants were recruited through their monthly drill exercise and each squad of two fire teams was randomly assigned to the virtual or live condition.

The population had a mean of 12.1 months of service in the ROTC, with a range of 4 to 24 months. Lastly, 6.25% of the population had prior room clearing training. This lack of deployments and training will define the population as "novice". The participants were provided with consent forms and given the opportunity to review the experiment objectives. Investigators were available to answer any questions. The participants will be asked to sign the consent forms to indicate they understand their participation is voluntary. The second study was conducted in the spring of 2015. The hypotheses tested during this study are discussed below.

Hypothesis 3: Performance Effect of Different Training Conditions of Individual Novice Soldiers

Hypothesis three explores the primary objective of this study and is posed to determine if there is a difference on the performance of novice soldiers who are trained using traditional means versus a game based virtual environment. It is hypothesized that "*for the task of clearing a room, there is no difference between the performances of novice soldiers who have been trained using* *traditional classroom means versus novice soldiers who have been trained using game based virtual environments*."

The independent variable for this hypothesis is the training condition and the dependent variable is the evaluation provided by the subject matter expert(s). The list of variables for $H3_0$ are shown in Table 5.

Table 5. List of Variables for Hypothesis 3

Variable	Type	Measurement Method	Measurement Scale
Training Condition	Independent	n/a	
Task Performance	Dependent	Subject Matter Expert Evaluation	Pass/Fail

Hypothesis 4: Performance Effect of Different Training Conditions of Teams of Novice Soldiers

Hypothesis four also explores the second objective of this study and is posed to determine if there is a difference on the performance of teams of novice soldiers who are trained using traditional means versus a virtual environment. It is hypothesized that "*for the task of clearing a room, there is no difference between the performances of teams of novice soldiers who have been trained using traditional means versus teams of novice soldiers who have been trained using game based virtual environments.*"

The independent variable for this hypothesis is the training condition and the dependent variables are the novice team's task performance and the relative rank of the team's performance. The task performance and ranks are provided by subject matter expert(s). The list of variables for $H4_0$ are shown in Table 6.

Table 6. List of Variables for Hypothesis 4

Variable	Type	Measurement Method	Measurement Scale
Training Condition	Independent	n/a	
Team Performance	Dependent	Subject Matter Expert Evaluation	Pass/Fail
Team Rank	Dependent	Subject Matter Expert Evaluation	Integer

Study #3: Population Comparisons

Hypothesis 5: Performance Comparison of Experienced Soldiers to Novice Soldiers

Hypothesis five addresses the third objective of these studies and will determine if there is a difference between the experienced and novice soldiers who are trained using the two different conditions. It is hypothesized that "*for the task of clearing a room, there is no difference between the performances of experienced soldiers and novice soldiers who have been trained using traditional means and game based virtual environments.*"

The independent variables for this hypothesis are the training condition and the soldier experience condition. The dependent variables include their task performance for each training condition with evaluation provided by the subject matter expert(s). The list of variables for $H5_0$ are shown in Table 7.

Table 7. List of Variables for Hypothesis 5

Variable	Type	Measurement Method	Measurement Scale
Training Condition	Independent	n/a	
Soldier Experience Condition	Independent	n/a	
Task Performance (Experienced Individuals)	Dependent	Subject Matter Expert Evaluation	Pass/Fail
Task Performance (Novice Individuals)	Dependent	Subject Matter Expert Evaluation	Pass/Fail
Task Performance (Experienced Teams)	Dependent	Subject Matter Expert Evaluation	Pass/Fail
Task Performance (Novice Teams)	Dependent	Subject Matter Expert Evaluation	Pass/Fail

Apparatus

For this experiment, a generic virtual world based on open source software was utilized. The decision to use this software was made deliberately so that no conclusions would be drawn based on an existing deployed product in the U.S. Army. Additionally, the two field experiments used in the data collection events are conducted at multiple locations.

Client Hardware

Although slightly different portable computers provided to the soldiers for each experiment, all computers were homogenous at each experiment. For example, all of the FLARNG 2/124th soldiers were provided Hewlett Packard model 8730w mobile workstations with the same version of the simulator software and the University of Central Florida U.S. Army ROTC cadets were provided with Hewlett Packard model 17-e118dx laptops. Table 8 outlines the specifics for client hardware deployment for both studies.

Table 8. Client Hardware Deployment

	Platform	CPU	RAM	GPU
Baseline VBS3 Specification	Intel or Compatible Workstation	Core 2 Duo 2.4 Ghz	4 Gb	Nvidia Quadro FX 3700m
Study #1: FLARNG 2/124th Leesburg, Florida	Hewlett Packard Elitebook 8730w	Intel Core 2 Duo P8600 (2.4Ghz)	4 Gb	Nvidia Quadro FX 3700m
Study #2: UCF ROTC Orlando, Florida	Hewlett Packard Pavilion 17-e118dx	AMD A8-4500M (2.8 Ghz)	8 Gb	AMD Radeon 7640G

Both hardware platforms meet or exceed the recommended minimum system requirements for the Virtual Battle Spaces 3 application, the U.S. Army's game for training program of record ("Virtual Battle Spaces 3 (VBS3)," 2010).

The primary means of operating the system was through keyboard and mouse. An external mouse and pad was provided to each soldier and they used the built-in keyboard for any typing. The soldiers were seated in close proximity to their team-mates and did not require microphone or audio equipment.

In both studies, the simulator was supported by a local area network with portable server equipment onsite.

Simulator Software

The simulator used for this study is the Military Open Simulator Enterprise Strategy (MOSES), a small research initiative supported by the U.S. Army Research Laboratory (Ortiz & Maxwell, 2014). The MOSES is composed of two major software components, a client and a server. The software is open source and has been modified for military research and development purposes.

The MOSES server software is derived from the Open Simulator project (Casey, 2014) and was chosen for a number of reasons. First, it is open source and is cost effectively modifiable to the needs of the user. Secondly, it can easily handle the demands of 50 or more simultaneous users out of the box (Gabrielova & Lopes, 2014).

The MOSES client software is derived from the open source Firestorm viewer (Lyon, 2014). The client's interface was modified so that the participants were not distracted or overwhelmed by unnecessary functionality. Only the controls or interfaces that were necessary to accomplishing the tasks presented were allowed to be shown to the user. This simplified user interface allowed for rapid acclimation and eased the transition from learning the software controls to the actual task training, Figure 5.

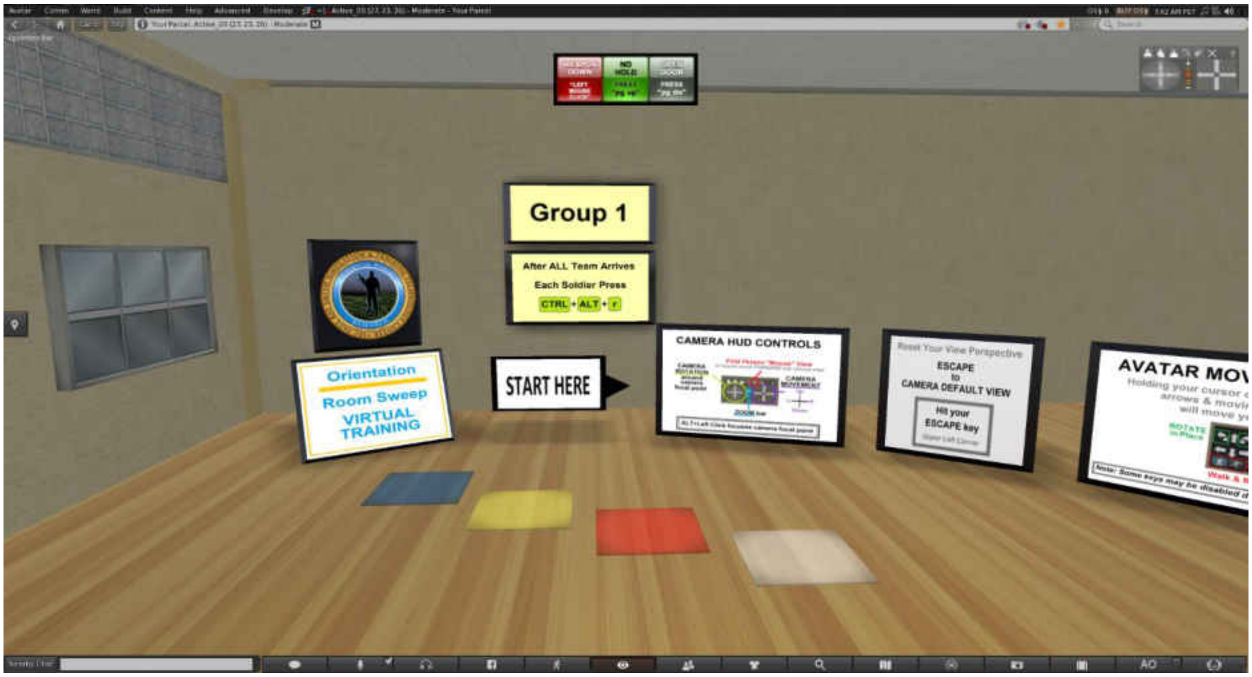

Figure 4. Simplified MOSES Interface

Experimental Procedures

The experimental procedure for Study #1 and Study #2 is identical for each population of soldiers. The soldiers assemble at the appointed time and place and fall into formation for the initial briefing from the project investigators. As a collective, they are told the participation in the study is voluntary and provided an overview of the project goals.

Study #1 Experimental Procedure

The 2/124th FLARNG soldiers (participants) were then randomly placed into two groups and asked to sit either at tables with workstations or proceed to a classroom (the UCF ROTC participants will receive similar treatment). Copies of the consent form are provided to all participants and sufficient time is allotted to allow for review of the evaluation objectives. Project investigators were available to answer any questions. The participants are then asked to sign the consent form, indicating they volunteer to participate. The participants are also asked for permission to be photographed and video recorded. The soldiers were told they may choose not to participate and that no reason will be given to their superiors, only that they did not meet project evaluation criteria.

Every participant completed a demographics questionnaire to provide data regarding military experience, training background, level of education, computer proficiency, and video game experience, Figure 5.

Figure 5. 2/124th Apache Co. Soldiers Completing Questionnaires

Study #2 Experimental Procedure

For the "Live" condition, the 2/1124th FLARNG soldiers in the classroom were provided with traditional lecture and presentation of room clearing task training. This material was derived from U.S. Army training doctrine such as FM 3-21.8 (*FM 3-21.8 The Infantry Rifle Platoon and Squad*, 2007). The instructor for the live condition was an experienced subject matter expert and possessed knowledge of the task, its execution in the field, and advanced knowledge of task completion.

For the "Virtual World" condition, the participants were seated at tables with laptops and are provided an external mouse. Each participants was provided with a workstation. The participants first watched a video that provided an overview of the room clearing task with material similarly derived from FM 3-21.8. The video was produced with machinima (footage obtained from the operation of the simulator) from within the virtual world simulator the participants used.

After the completion of the video, the soldiers were then allowed 30 minutes of acclimation time within the simulator. The simulator training consisted of a series of small tasks the soldiers were asked to perform while following a trail.

At the end of the virtual task trail, the participants were presented with an obstacle course. To complete the obstacle course, the participants must use all of the skills presented in the task trail. By completing the obstacle course and arriving at the staging area, they demonstrated proficiency in the simulator and are ready to proceed with the room clearing training. The simulator user interface was carefully created to allow the participant to only perform actions or tasks required, thus reducing the time for acclimation.

Participants were then allowed five training trials within the virtual world. Each of the trials required the participants to complete the room clearing task as part of a fire team. The views presented to the participants within the virtual world are tuned to the role of the soldier in the fire team. This created a situation such that the client software was made to be as simple as possible to use, reduced the amount of time required to train the participant in the use of the software, and allowed for only the precise amount of information required by the participant to do the job was presented.

After completion of the training, the participants undergo a final assessment in a live environment. The experimenter informed the participants they will perform the same room clearing task in the live environment that they encountered in either of the training conditions. The participants were allowed up to two assessed attempts as part of a Fire Team within a real room.

After the completion of the live assessment, all participants were provided feedback from the subject matter expert for both individual and team performance. If the team's performance was unsatisfactory during the first live assessment, they were allowed one more attempt. Lastly, the participants were provided with a copy of the consent form.

Data Analysis Methods

This dissertation is composed of three studies. Study #1 uses experienced soldiers and #2 uses novice soldiers, both using a single-factor completely randomized design. The soldiers in each study were randomly assigned to one of two training conditions, the live treatment and the virtual treatment. Study #3 uses combined data collected from #1 and #2. For Study #1 and #2, the independent variable is the training condition (live or virtual) and the dependent variables are task performance and team rank. Data collection will be performed systematically and these are Single Blind Tests for the subject matter expert Evaluators. Descriptive statistics for all data collected, including means and standard deviations, will be calculated. Since some of the data will not be normalized, specifically the number of teams for the two soldier experience conditions, non-parametric tests will be performed to analyze Hypothesis 2 and 4. An alpha of 0.05 will be used for all significance tests. Exact p-values will be reported unless the probability falls below 0.001.

There are 64 individual experienced soldiers in Study #1, therefore normalized statistics may be used for Hypothesis 1. Since the performance data is categorical, a chi-square analysis is performed to determine whether the individual training conditions are dependent or independent. A logistic regression will also be performed to determine probability of one treatment is more or less likely to obtain a passing score. These analyses will be performed for each round of assessment.

For Hypothesis 2, there are only 16 teams of experienced soldiers available in Study #1, non-parametric statistics will be used. The subject matter experts will provide both performance data and ranking data in the assessment. This data will be used to perform a Mann Whitney T Test for performance (pass/fail) by training condition. This T Test will be performed to determine if one training condition yields better performance than another by teams of experienced soldiers.

Since there are 64 individual novice soldiers in Study #2, normalized statistics may be used to analyze data for Hypothesis 3. The performance data for this assessment is also categorical (Go/No-Go), a chi-square analysis is performed to determine whether the individual training conditions are dependent or independent. A logistic regression will also be performed to determine probability of one treatment is more or less likely to obtain a passing score. These analyses will be performed for each round of assessment.

For Hypothesis 4, there are only 16 teams of novice soldiers available in Study #2, non-parametric statistics will be used. The subject matter experts will provide both performance data and ranking data in the assessment. This data will be used to perform a Wilcoxon T test for performance (pass/fail) by training condition. This T Test will be performed to determine if one training condition yields better performance than another by teams of novice soldiers.

For Hypothesis 5, normalized statistics will be used since both individual and team quantities are sufficiently high. This analysis will look for statistically relevant differences in performance by the experienced or novice soldiers when comparing the performances from the training conditions. The first test will be a comparison of all individual performances for the baseline training treatment versus the virtual treatment for each round of assessment. The next test will be a comparison of individual experienced to novice soldier performances for the baseline treatment for each round of assessment. The next test will be a comparison of individual experienced to novice soldier performance for the virtual treatment for each round of assessment. For the collective cases, a test is performed for a comparison of all soldier's team's performances for the baseline versus virtual treatments.

Lastly, DSSQ data will be analyzed using a 2 by 2 between-groups multi-variance analysis to assess the effect of the two training treatments on subject distress, engagement, and worry. A discussion of this data and how it may assist in the explanation for the outcomes of the data analysis for the five hypothesis is provided in chapter five. A summary is provided in Table 9.

Table 9. Performance Data Analysis Methods

Study		Method	n	Variables
#1	$H1_0$	• Chi-square Analysis • Logistic Regression	64	• Training Condition (Live/Virtual) • Individual Task Performance (Pass/Fail)
	$H2_0$	• Non-Parametric Mann-Whitney Test	16	• Training Condition • Task Performance • Performance Ranking (Relative)
#2	$H3_0$	• Chi-square Analysis • Logistic Regression	64	• Training Condition • Task Performance
	$H4_0$	• Non-Parametric Mann-Whitney Test	16	• Training Condition • Task Performance • Performance Ranking
#3	$H5_0$	• Chi-square Analysis • Logistic Regression	128 32	• Training Condition • Task Performance
DSSQ		• MANOVA	64	• Live vs Virtual Between Subjects • Pre vs Post Training Within Subjects

CHAPTER FOUR: DATA AND ANALYSIS

Chapter Four Summary

Chapter four is a discussion of the data analysis and findings of the studies discussed in chapter 3. Two populations of soldiers were used in these studies, novice and experienced. They were also provided two training treatments, traditional classroom and simulation based virtual simulators. Additionally, their performance was assessed individually as well as a collective. This chapter contains a discussion of hypotheses test results and an in-depth interpretation of the empirical data.

The findings Study #1 include both the total data collected as well as an analysis of data with outliers removed. Study #1 examined performance of experienced soldiers in both an individual and collective manner (Hypothesis #1 and #2). Similarly, Study #2 examined performance of novice soldiers in a individual and collective manner (Hypothesis #3 and #4) and this data also had outlier data. Lastly, Study #3 looked at the combined performance data of both experienced and novice groups.

Study #1: Experienced Individual Performance Findings

The soldier population used in study #1 was 64 male reservists, composed of 16 fire teams of four soldiers each. The participants ranged in age from 19 to 35 years with a mean of 25.22 and standard deviation of 3.8. This population has a mean of 4.49 years of service, with a range of 0.58 to 11 years. This population is considered "experienced" as defined by ranged in number of deployments from none to three, with a mean of 0.69 and 94% of the population had prior room clearing training.

The experienced group's individual performance was assessed subjectively by the subject matter expert rater at the time of task completion. Although the task of clearing a room is performed as a collective, each position in the team is unique and can be assessed individually. The assessment is provided as a "GO or NO-GO" rating, which indicated whether the soldier completed their task to the SME's satisfaction or not. The SMEs rating the individual and collective performance followed the 4 step rubric discussed in Chapter 3. The ratings included the entry phase, the eliminate threat phase, attaining a position of dominance, and consolidation and reorganization.

A post hoc power analysis was performed using a chi-square goodness of fit. Using a medium effect according to Cohen's Convention (Cohen, 1992), a power of 0.67 was calculated for the given sample size of 64 soldiers.

Table 10 reports the means and standard deviations for the room clearing performance of the experienced population. After the first trial, "Round 1", 72% of the soldiers provided with the baseline treatment were given a "Go" rating and 50% of the soldiers provided with the virtual treatment were given a "Go" rating, as reported in Table 10. After the second trial, "Round 2", 84% of the baseline treatment passed and 100% of the virtual treatment passed.

Table 10. Overall Performance Means and Standard Deviations (Experienced Group)

		Round 1		Round 2	
		Baseline	Virtual	Baseline	Virtual
Experienced	Mean	0.72	0.5	0.84	1
	SD	0.456	0.508	0.368	0

Figure 6 visually depicts the dependent variable, Performance, by training treatment and by trial.

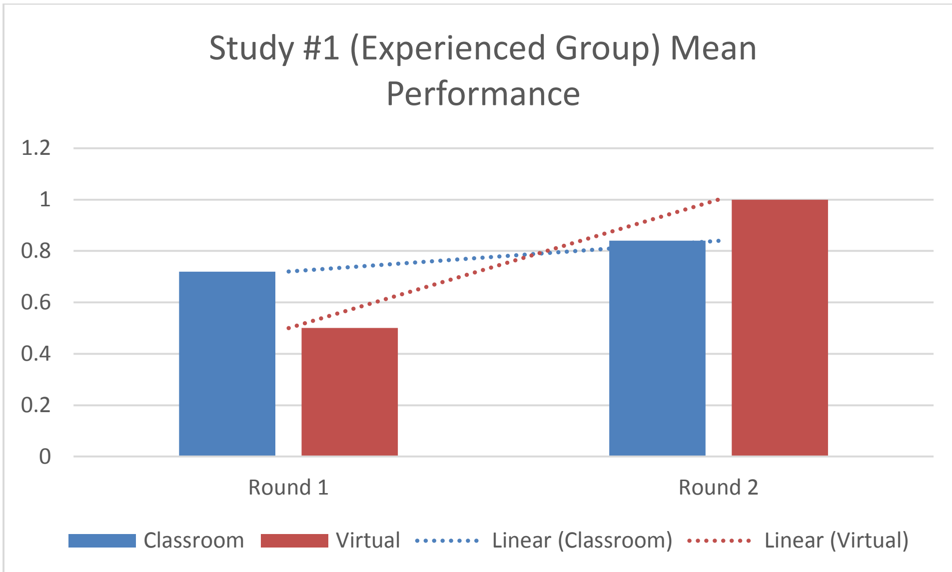

Figure 6. Study #1 Experienced Mean Performance

Hypothesis 1 Test Results

Hypothesis 1 was "*for the task of clearing a room, there is no difference between performances of experienced soldiers who have been trained using traditional classroom means versus experienced soldiers who have been trained using game based virtual environments.*" Table 11 presents the performance results of the individuals after the first round of live assessment. The Chi-square analysis for the Round #1 assessment data shows a probability of 0.07, where$\chi^2(1, n = 64) = 3.22$. Since this is more than our alpha of 0.05, the analysis indicate the two conditions are independent and the hypothesis is supported.

Table 11. Individual Experienced Soldier Performance after Round #1 Assessment

	Go	No-Go	Total
Baseline	23	9	32
Virtual	16	16	32
Total	39	25	64

A logistic regression shows an odds ratio of 39%. This means the regression indicates that the soldiers trained with the virtual treatment were less likely to obtain a "Go" rating in the live assessment.

Table 12. Logistic Regression for Experienced Soldier Performance after Round #1 Assessment.

	StdErr	Odds Ratio
Virtual Treatment	0.5288	0.39

Table 13 presents the performance results of the individuals after the second round of live assessment. The chi-square analysis shows a probability of 0.0067, where $\chi^2(1, n = 64) = 7.3557$. The result is significant and is well below the alpha of 0.05, indicating the two conditions are dependent and the hypothesis is now *not* supported.

Table 13. Individual Experienced Soldier Performance after Round #2 Assessment

	Go	No-Go	Total
Baseline	27	5	32
Virtual	32	0	32
Total	59	5	64

Hypothesis 1 Test Results with Outlier Data Removed

An examination of the data revealed that four soldiers had no prior room clearing training and no deployments, two of those soldiers had 1 year or less in service. Of the five soldiers who did not pass the second round of assessment, three of the five had no deployments and had the rank of Private First Class. A second analysis was performed, removing the performance data for the four inexperienced solders. Table 14 provides a report of the adjusted means and standard deviations for the room clearing performance of the experienced population with the outliers removed. In this data, 100% of the soldiers had prior room clearing training. In this adjusted data, 79% of the baseline treatment received a "Go" rating after round 1 assessment and 89% of the baseline treatment received a "Go" rating after round 2 assessment. The virtual treatment remained unchanged at 50% after round 1 and 100% after round 2 assessment.

Table 14. Overall Performance Means and Standard Deviations (Experienced Group, Outliers Removed)

		Round 1		Round 2	
		Baseline	**Virtual**	**Baseline**	**Virtual**
Experienced	**Mean**	0.79	0.5	0.89	1
	SD	0.417	0.508	0.315	0

Table 15 provides a summary of the adjusted data with the outliers removed. The chi-square analysis for round 1 performance shows a probability of 0.022, where $\chi^2(1, n = 60) = 5.25$. The result is significant and now indicates the hypothesis is no longer supported. The chi-square analysis for round 2 performance shows a probability of 0.057, where $\chi^2(1, n = 60) = 3.61$. This result is not significant and also indicates the hypothesis is now supported.

Table 15. Adjusted Experienced Performance Ratings with Outliers Removed

Round 1	Go	No-Go	Total
Baseline	22	6	28
Virtual	16	16	32
Total	38	22	60
Round 2	**Go**	**No-Go**	**Total**
Baseline	25	3	28
Virtual	32	0	32
Total	57	3	60

Figure 7 visually depicts the adjusted dependent variable with outliers removed, Performance, by training treatment and trial.

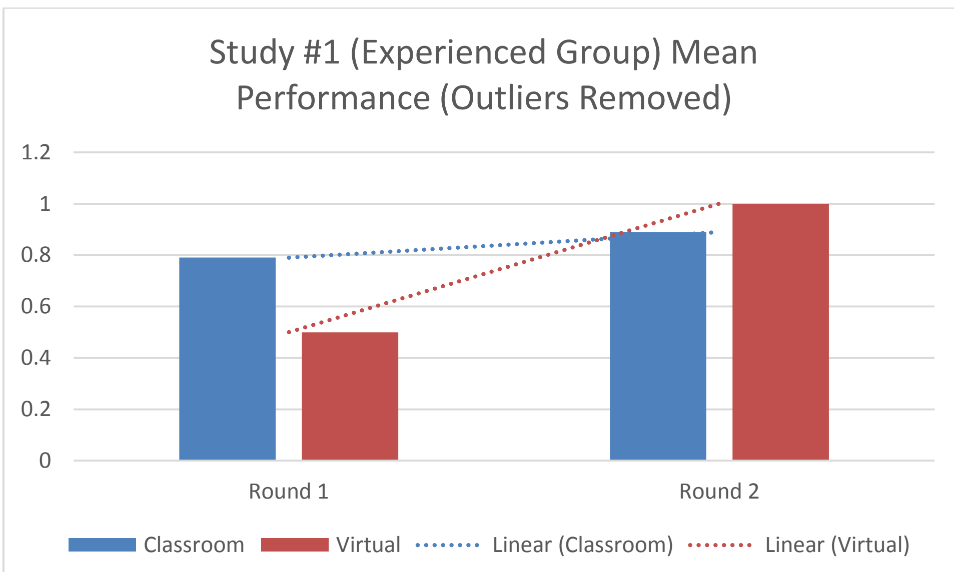

Figure 7. Study #1 Experienced Mean Performance with Outliers Removed

Study #1 Experienced Collective Performance Findings

The experienced group's collective training event utilized onsite facilities at the 2/124th FLARNG Leesburg Armory. These activities were video recorded so that a subject matter expert could later evaluate the team's performance and provide a relative ranking. Where appropriate, the soldiers were allowed to remain in their organic fire teams.

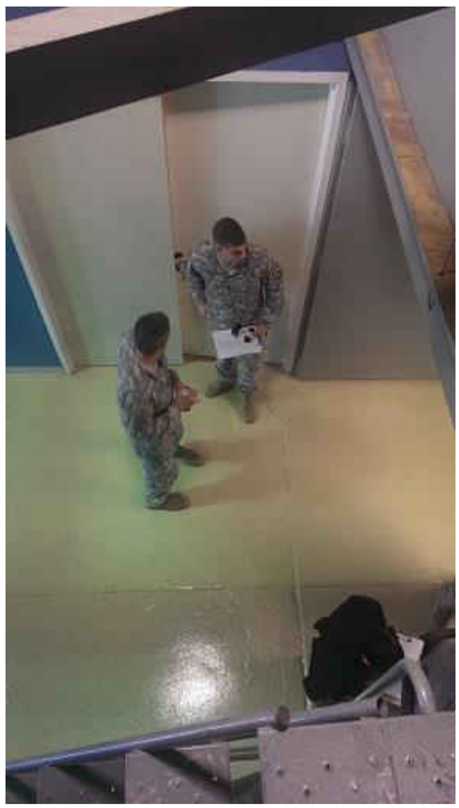

Figure 8. 2/124th FLARNG Training Evaluators

Hypothesis 2 Test Results

Hypothesis 2 was "*for the task of clearing a room, there is no difference between the performances of teams of experienced soldiers who have been trained using traditional means versus teams of experienced soldiers who have been trained using game based virtual environments.*" Table 16 shows the ratings and relative rankings of the teams for the experienced group. The subject matter expert subjectively reviewed the performance of the teams and provided a relative ranking for each.

Table 16. Collective Soldier Performance (Experienced Group)

Treatment	Team	Rating	Relative Ranking
Traditional	1	0	9
	2	1	7
	3	N/A	N/A
	4	0	10.5
	5	0	13
	6	1	3
	7	0	12
	8	1	4.5
Virtual	9	1	1
	10	1	2
	11	0	10.5
	12	1	7
	13	1	7
	14	N/A	N/A
	15	0	14
	16	1	4.5

The nonparametric test chosen to analyze the null hypothesis is the Mann-Whitney U test. The data was ranked by the subject matter expert in an ordinal fashion, given that there are 16 total teams they were ranked by performance on a scale from 1 to 16. All observations are independent of each other. A situation arose in each treatment that one team could not be rated or ranked, therefor 7 teams from each treatment were ultimately rated and ranked.

A Mann-Whitney test indicated no significant difference in performance between the virtual treatment (Mdn = 9) and baseline treatment (Mdn = 7), $U = 18$, $p = 0.44$. Since this result is higher than 0.05, the hypothesis is supported and conclude there is no difference in performance.

Study# 2 Novice Individual Performance Findings

The population used in Study #2 was 64 ROTC cadets, composed of 16 fire teams. The participants ranged in age from 18 to 28 years with a mean of 19.9 and a standard deviation of 1.84. There were 47 male and 17[1] female participants. This population is considered "novice" as defined by a mean of 12.1 months in the UCF ROTC, with a range of 4 to 24 months of service. 6.25% of this population had prior room clearing or building clearing experience. Due to time constraints, one cadet did not complete the live assessment.

Figure 7 visually depicts the dependent variable, Performance, by training treatment and by trial. After the first trial, "Round 1", 47% of the soldiers provided with the baseline treatment were given a "Go" rating and 55% of the soldiers provided with the virtual treatment were given a "Go" rating. After the second trial, "Round 2", 78% of the baseline treatment passed and 78% of the virtual treatment passed.

---

[1] 64 Participants participated in the training portion of the experimentation, however one female participant left the live evaluation phase early. Data from 63 ROTC cadets were used in the performance analysis.

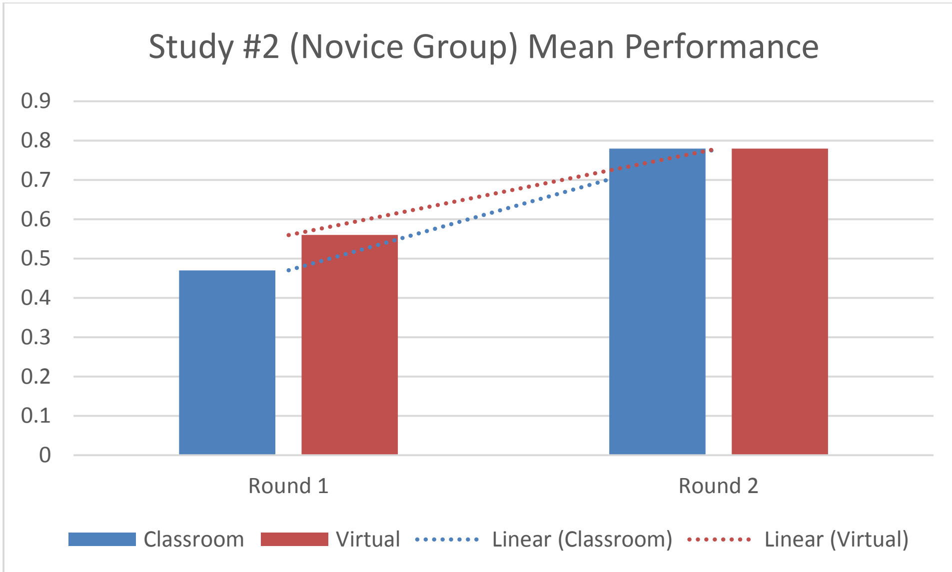

Figure 9. Study #2 Novice Mean Performance

Table 17 reports the means and standard deviations for the room clearing performance of the novice population.

Table 17. Overall Performance Means and Standard Deviations (Novice Group)

		Round 1		Round 2	
		Baseline	**Virtual**	**Baseline**	**Virtual**
Novice	**Mean**	0.47	0.55	0.78	0.774
	SD	0.507	0.506	0.42	0.425

Hypothesis 3 Test Results

Hypothesis 3 was "*for the task of clearing a room, there is no difference between the performances of novice soldiers who have been trained using traditional classroom means versus novice soldiers who have been trained using game based virtual environments*". Table 18 presents the performance results of the individuals after the first round of live assessment. The chi-square analysis shows a probability of 0.3996, where$\chi^2(1, n = 63) = 0.527$. Since this is more than our alpha of 0.05, the analysis indicate the two conditions are independent and the hypothesis is supported.

Table 18. Individual Novice Soldier Performance after Round #1 Assessment

	Go	No-Go	Total
Baseline	15	17	32
Virtual	17	14	31
Total	32	31	63

A logistic regression shows an odds ratio of 1.37, Table 19. This means the regression indicates that the soldiers trained with the virtual treatment have 137% greater odds to obtain a "Go" rating in the live assessment.

Table 19. Logistic Regression for Novice Soldier Performance after Round #1 Assessment.

	StdErr	Odds Ratio
Virtual Treatment	0.5057	1.37

Table 20 presents the performance results of the individuals after the second round of live assessment. Those soldiers who did not pass the first round were allowed to try again. The chi-square analysis shows a probability of 0.946, where$\chi^2(1, n = 63) = 0.0045$. This result is not significant and is well above the alpha of 0.05, indicating the two conditions are not dependent and the hypothesis is supported.

Table 20. Individual Novice Soldier Performance after Round #2 Assessment

	Go	No-Go	Total
Baseline	25	7	32
Virtual	24	7	31
Total	49	14	63

A logistic regression shows an odds ratio of 0.96, Table 21. This means the regression indicates that the soldiers trained with the virtual treatment were slightly less likely to obtain a "Go" rating in the live assessment.

Table 21. Logistic Regression for Novice Soldier Performance after Round #2 Assessment.

	StdErr	Odds Ratio
Virtual Treatment	0.6061	0.96

Hypothesis 3 Test Results with Outlier Data Removed

According to the UCF ROTC demographics data, four of the participants had previous room clearing experience. An additional analysis was performed excluding the performance data of the four experienced participants. In this data, 0% of the soldiers had prior room clearing training. In this adjusted data, 48.4% of the baseline treatment received a "Go" rating after round 1 assessment and 77.4% of the baseline treatment received a "Go" rating after round 2 assessment. 58.6% of the participants with the virtual treatment passed after round 1 and 82.7% after round 2 assessment. Table 22 reports the means and standard deviations for the room clearing performance of the novice population with outliers removed. Table 23 reports a summary of the adjusted round 1 and round 2 performance ratings with the outliers removed.

Table 22. Overall Performance Means and Standard Deviations (Novice Group, Outliers Removed)

		Round 1		Round 2	
		Baseline	Virtual	Baseline	Virtual
Novice	Mean	0.48	0.59	0.77	0.83
	SD	0.507	0.501	0.425	0.384

Table 23. Adjusted Novice Performance Ratings with Outliers Removed

Round 1	Go	No-Go	Total
Baseline	15	16	31
Virtual	17	12	29
Total	32	28	60
Round 2	**Go**	**No-Go**	**Total**
Baseline	24	7	31
Virtual	24	5	29
Total	48	12	60

The chi-square analysis for round 1 performance shows a probability of 0.427, where $\chi^2(1, n = 60) = 0.6305$. The result is not significant and indicates the hypothesis is supported. The chi-square analysis for round 2 performance shows a probability of 0.6050, where $\chi^2(1, n = 60) = 0.267$. This result is not significant and also indicates the hypothesis is now supported. Figure 10 visually depicts the adjusted dependent variable with outliers removed, Performance, by training treatment and trial.

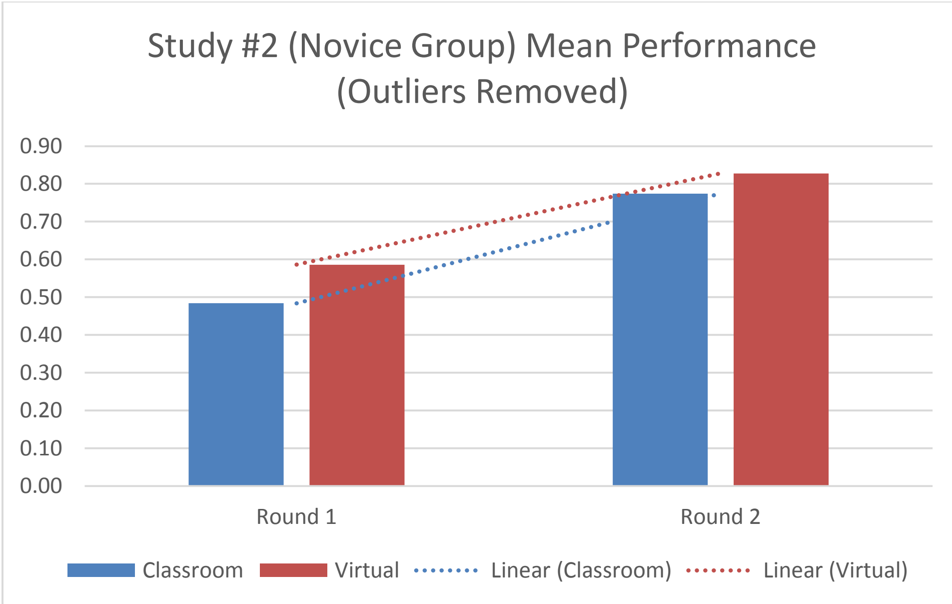

Figure 10. Study #2 Novice Mean Performance with Outliers Removed

Study #2 Novice Collective Performance Findings

The novice group's collective training event utilized onsite facilities at the UCF Campus ROTC Battle Lab and multipurpose room. These activities were video recorded so that a subject matter expert could later evaluate the team's performance and provide a relative ranking. Where appropriate, the soldiers were allowed to remain in their organic fire teams.

Hypothesis 4 Test Results

Hypothesis 4 was "*for the task of clearing a room, there is no difference between the performances of teams of novice soldiers who have been trained using traditional means versus teams of novice soldiers who have been trained using game based virtual environments.*" Table 24 provides a summary of the novice group collective performance ratings and relative rankings.

Table 24. Collective Soldier Performance (Novice Group)

Treatment	Team	Rating	Relative Ranking
Virtual	1	0	11
	2	1	4
	3	1	6.5
	4	1	8.5
	5	0	15
	6	1	10
	7	1	3
	8	1	1
Traditional	9	1	6.5
	10	1	2
	11	1	8.5
	12	1	12
	13	1	5
	14	1	14
	15	0	16
	16	1	13

A Mann-Whitney test indicated no significant difference in performance between the Live (*Mdn* = 10.3) and Virtual (*Mdn* = 7.5) treatments, $U = 23$, $p = 0.37$. Since this result is higher than 0.05, the hypothesis is supported and conclude there is no difference in performance.

Study #3 Population Comparison Findings

Hypothesis five states that "*for the task of clearing a room, there is no difference between the performances of experienced soldiers and novice soldiers who have been trained using traditional means and game based virtual environments.*" To study this hypothesis, a series of chi-square tests were calculated against combinations of pairs of interactions. These combinations were explored after round 1 and round 2 of the performance assessments. Since this is an examination of data from both experienced and novice populations, outliers were not removed from either population.

Hypothesis 5 Test Results: Individual Performance Comparisons

The individual performances were analyzed by treatment and by experience level. Table 25 provides a summary of all soldier's performance after round 1 assessment. 59% of the 64 soldiers trained using traditional baseline means passed their assessment on the first try, versus 52% of the 63 soldier trained using virtual means receiving a "Go" rating on the first try. The Chi-square analysis of this data shows a probability of 0.427, where$\chi^2(1, n = 127) = 0.63$. Since this probability is higher than our alpha, the analysis indicates the two conditions are independent and the hypothesis is supported.

Table 25. Comparison of All Soldier's Individual Performance for the Baseline versus Virtual Treatments after Round 1 Assessment

Round 1	Go	No-Go	Total
Baseline	38	26	64
Virtual	33	30	63
Total	71	56	127

Table 26 provides a summary of all soldier's performance after round 2 assessment. 81% of the soldiers trained using traditional baseline means passed their assessment on the second try, versus 89% of the soldier trained using virtual means receiving a “Go” rating on the second try. The Chi-square analysis of this data shows a probability of 0.227, where$\chi^2(1, n = 127) = 1.456$. Since this probability is higher than our alpha, the analysis indicates the two conditions are independent and the hypothesis is supported.

Table 26. Comparison of All Soldier’s Performance for the Baseline versus Virtual Treatments after Round 2 Assessment

Round 2	Go	No-Go	Total
Baseline	52	12	64
Virtual	56	7	63
Total	108	19	127

The next combination tested was a comparison of the performance of individual experienced soldiers to novice soldiers who were provided baseline based training after round 1 assessment. The chi-square analysis shows a probability of 0.041739, where$\chi^2(1, n = 64) = 4.1457$. This result is slightly below the threshold of alpha of 0.05 and the hypothesis is not supported and there is an indication of an advantage for the experienced group. Table 27 provides the summary of individual experienced to individual novice soldier’s round 1 assessments.

Table 27. Comparison of Individual Experienced to Novice Soldier Performance for the Baseline Treatment after Round 1 Assessment

Round 1 Baseline	Go	No-Go	Total
Experienced	23	9	32
Novice	15	17	32
Total	38	26	64

Table 28 provides a summary of individual experienced to novice soldier performances who were provided baseline training after round 2 assessment. The chi-square analysis shows a probability of 0.5218, where $\chi^2(1, n = 64) = 0.4103$. Since this probability is higher than our alpha, the analysis indicates the two conditions are independent and the hypothesis is supported.

Table 28. Comparison of Individual Experienced to Novice Soldier Performance for the Baseline Treatment after Round 2 Assessment

Round 2 Baseline	Go	No-Go	Total
Experienced	27	5	32
Novice	25	7	32
Total	52	12	64

Table 29 provides a round 1 performance summary of individual experienced and novice soldiers who were provided the virtual training treatment. The chi-square analysis shows a probability of 0.7, where $\chi^2(1, n = 64) = 0.1478$. Since this probability is higher than our alpha, the analysis indicates the two conditions are independent and the hypothesis is supported.

Table 29. Comparison of Individual Experienced to Novice Soldier Performance for the Virtual Treatment after Round 1 Assessment

Round 1 Virtual	Go	No-Go	Total
Experienced	16	16	32
Novice	17	14	31
Total	33	30	63

Table 30 provides the round 2 performance summary of individual experienced and novice soldiers who were provided virtual training. The chi-square analysis shows a very small probability value of 0.004, where $\chi^2(1, n = 63) = 8.129$. This probability is nearly zero, thus the analysis indicates the hypothesis is not supported in this case.

Table 30. Comparison of Individual Experienced to Novice Soldier Performance for the Virtual Treatment after Round 2 Assessment

Round 2 Virtual	Go	No-Go	Total
Experienced	32	0	32
Novice	24	7	31
Total	56	7	63

Hypothesis 5 Test Results: Collective Performance Comparison

The collective performances were analyzed by training treatment and experience. 80% of the teams provided with baseline training were given a "Go" rating and 60% of the teams provided with virtual training were given a "Go" rating. Table 31 provides a summary of the soldier's performance comparisons by treatment. The Chi-square analysis shows a probability of 0.23, where$\chi^2(1, n = 30) = 1.428$. Since this probability is higher than our alpha, the analysis indicates the two conditions are independent and the hypothesis is supported.

Table 31. Comparison of All Soldiers' Collective Performance for the Baseline versus Virtual Treatments

	Go	No-Go	Total
Baseline	12	3	15
Virtual	9	6	15
Total	21	9	30

A logistic regression shows an odds ratio of 0.375 at a 95% confidence interval. This means the regression indicates that the novice soldiers trained with the virtual treatment are less likely (37.5%) to obtain a "Go" rating in the live assessment, Table 32.

Table 32. Logistic Regression for Novice Soldier Performance after Round #1 Assessment.

	StdErr	Odds Ratio
Virtual Treatment	0.8333	0.375

Stress Questionnaire: Experienced Population

The subjective data used in this study was collected using self-reporting surveys which included stress. The DSSQ (Matthews et al., 2006, 2013) was used to collect stress data. An overview of both questionnaires was performed in chapter 3. In this analysis the experienced population was examined for effects of performance from distress, engagement, and worry.

Table 33 provides the descriptive statistics for the DSSQ pre distress and post distress means and standard deviations for the baseline and virtual treatments, standard error, and 95% confidence intervals. The data, for the baseline group, shows a mean score for pre-training Distress was 5.97 (SD=5.227) and the mean score for the post-training Distress was 6.06 (SD=5.599). The data for the soldiers who received the virtual treatment shows a mean score for pre-training Distress was 4.66 (SD=4.639) and the mean score for the post training Distress was 3.94 (SD=4.765).

Table 33. Descriptive Statistics for Pre and Post Distress Scores by Training Condition

	Condition	Mean	Std. Dev.	N	Std. Error	95% CI	
						Lower	Upper
DSSQ Pre Distress	Virtual	4.66	4.639	32	0.874	2.910	6.403
	Baseline	5.97	5.227	32	0.874	4.222	7.715
	Total	5.31	4.947	64			
DSSQ Post Distress	Virtual	3.94	4.765	32	0.919	2.100	5.775
	Baseline	6.06	5.599	32	0.919	4.225	7.900
	Total	5.00	5.267	64			

A two-way factorial analysis of variance within subject's DSSQ distress scores for virtual and baseline treatments was performed between pre and post-training data. A 2 by 2 between-groups multi-variance analysis was conducted to assess the effect of two training treatments on subject distress. The independent variables were the baseline training treatment and the virtual treatment. The dependent variables were the pre and post distress scores from the DSSQ. There was no significant interaction (α = 0.05) between the training conditions and pre-training distress, $F(1,62) = 1.129$, $p = 0.292$ or for post-training distress, $F(1,62) = 2.673$, $p = 0.107$. These results indicate there is no effect of the training condition on distress (Lackey, Salcedo, Matthews, & Maxwell, 2014).

Table 34 provides the descriptive statistics for the DSSQ pre-training engagement and post-training engagement means and standard deviations for the baseline and virtual treatments, standard error, and 95% confidence intervals. The data, for the baseline group, shows a mean score for pre-training Engagement was 23.47 (SD=5.594) and the mean score for the post-training Engagement was 22.69 (SD=6.488). The data for the soldiers who received the virtual treatment shows a mean score for pre-training Engagement was 24.47 (SD=4.697) and the mean score for the post training Engagement was 25.69 (SD=6.051).

Table 34. Descriptive Statistics for Pre and Post Engagement Scores by Training Condition

	Condition	Mean	Std. Dev.	N	Std. Error	95% CI	
						Lower	Upper
DSSQ Pre-Training Engagement	Virtual Baseline Total	24.47 23.47 23.97	4.697 5.594 5.148	32 32 64	0.913 0.913	22.644 21.644	26.294 25.294
DSSQ Post-Training Engagement	Virtual Baseline Total	25.69 22.69 24.19	6.051 6.488 6.404	32 32 64	1.109 1.109	23.471 20.471	27.904 24.904

A 2 by 2 between-groups multi-variance analysis was conducted to assess the effect of two training treatments on subject engagement. The independent variables were the baseline training treatment and the virtual treatment. The dependent variables were the pre and post engagement scores from the DSSQ. There was no significant interaction (α = 0.05) between the training conditions and pre-training engagement, $F(1,62) = 0.600$, $p = 0.442$ or for post-training engagement, $F(1,62) = 3.659$, $p = 0.060$. These results indicate there is no interaction of training condition and engagement.

Table 35 provides the descriptive statistics for the DSSQ pre-training worry and post-training worry means and standard deviations for the baseline and virtual treatments, standard error, and 95% confidence intervals. The data, for the baseline group, shows a mean score for pre-training Worry was 11.19 (SD=4.673) and the mean score for the post-training Worry was 11.09 (SD=6.280). The data for the soldiers who received the virtual treatment shows a mean score for pre-training Worry was 12.00 (SD=5.061) and the mean score for the post training Worry was 9.69 (SD=5.032).

Table 35. Descriptive Statistics for Pre and Post Worry Scores by Training Condition

	Condition	Mean	Std. Dev.	N	Std. Error	95% CI	
						Lower	Upper
DSSQ Pre-Training Engagement	Virtual	12.00	5.061	32	0.861	10.279	13.721
	Baseline	11.19	4.673	32	0.861	9.466	12.909
	Total	11.59	4.849	64			
DSSQ Post-Training Engagement	Virtual	9.69	5.032	32	1.006	7.677	11.698
	Baseline	11.09	6.280	32	1.006	9.083	13.105
	Total	10.39	5.689	64			

A 2 by 2 between -groups multi-variance analysis was conducted to assess the effect of two training treatments on subject worry. The independent variables were the baseline training treatment and the virtual treatment. The dependent variables were the pre and post worry scores from the DSSQ. There was no significant interaction (α = 0.05) between the training conditions and pre-training worry, $F(1, 62) = 0.445$, $p = 0.507$ or for post-training engagement, $F(1, 62) = 0.977$, $p = 0.327$. These results indicate there is no interaction of training condition and worry.

Figure 11 provides a consolidated summary of means for the DSSQ Distress, Engagement, and Worry scores for the experienced soldier population.

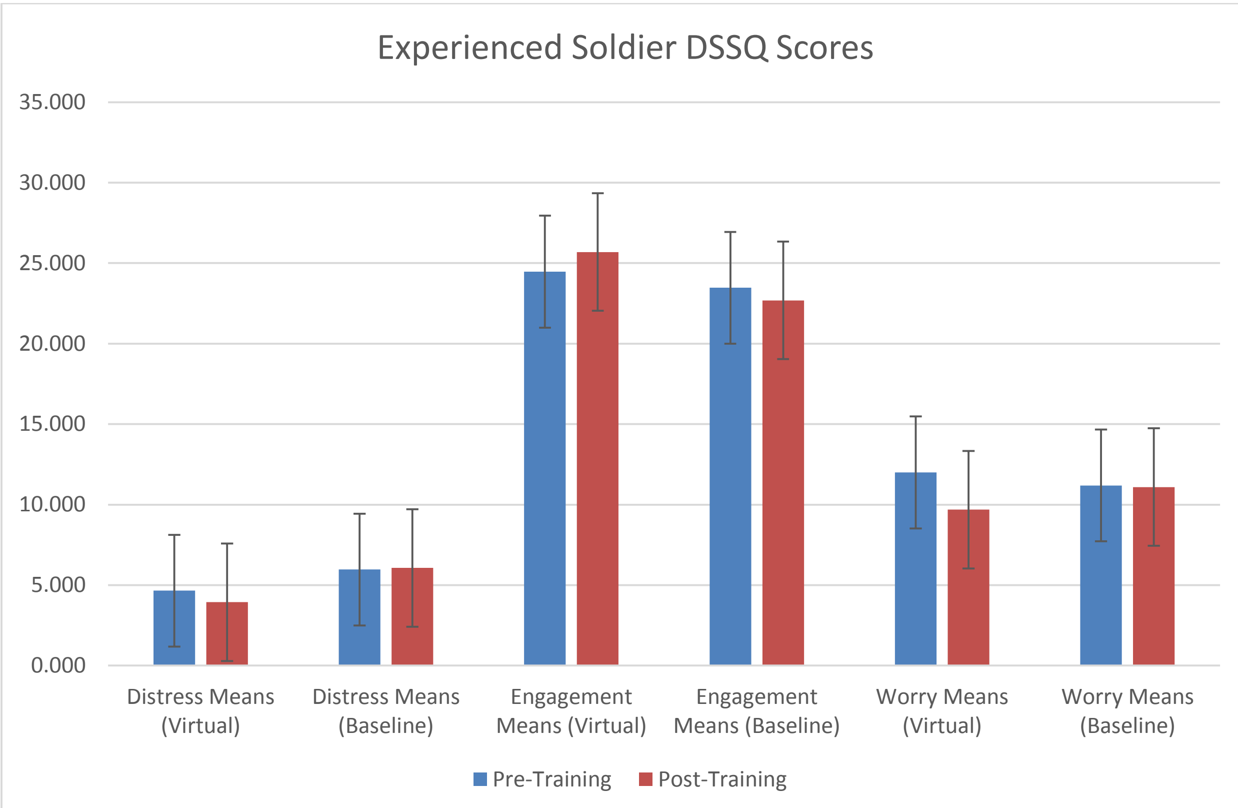

Figure 11. Summary of DSSQ Scores for Experienced Population

Stress Questionnaire: Novice Population

Table 36 provides the descriptive statistics for the DSSQ pre distress and post distress means and standard deviations for the baseline and virtual treatments, standard error, and 95% confidence intervals. The data, for the baseline group, shows a mean score for pre-training Distress was 5.28 (SD=4.199) and the mean score for the post-training Distress was 5.44 (SD=4.435). The data for the soldiers who received the virtual treatment shows a mean score for pre-training Distress was 8.03 (SD=5.894) and the mean score for the post training Distress was 6.53 (SD=5.913).

Table 36. Descriptive Statistics for Pre and Post Distress Scores by Training Condition

	Condition	Mean	Std. Dev.	N	Std. Error	95% CI	
						Lower	Upper
DSSQ Pre-Training Distress	Virtual Baseline Total	8.03 5.28 6.66	5.894 4.199 5.262	32 32 64	0.905 0.905	6.223 3.473	9.839 7.089
DSSQ Post-Training Distress	Virtual Baseline Total	6.53 5.44 5.98	5.913 4.435 5.214	32 32 64	0.924 0.924	4.684 3.590	8.378 7.285

A 2 by 2 between-groups multi-variance analysis was conducted to assess the effect of two training treatments on subject distress. The independent variables were the baseline training treatment and the virtual treatment. The dependent variables were the pre and post distress scores from the DSSQ. There was significant interaction ($\alpha = 0.05$) between the training conditions and pre-training distress, $F(1, 62) = 4.621$, $p = 0.035$ but not for post-training distress, $F(1, 62) = 0.701$, $p = 0.406$. The data shows elevated distress in the novice soldiers who received the virtual treatment. These results indicate there is effect of the training condition on distress for soldiers with no room clearing experience. Since the post-training condition shows no significant interaction, there is also an indication the virtual training could have alleviated the distress in some way.

Table 37 provides the descriptive statistics for the DSSQ pre-training engagement and post-training engagement means and standard deviations for the baseline and virtual treatments, standard error, and 95% confidence intervals. The data, for the baseline group, shows a mean score for pre-training Engagement was 27.31 (SD=3.995) and the mean score for the post-training Engagement was 29.16 (SD=3.743). The data for the soldiers who received the virtual treatment shows a mean score for pre-training Engagement was 22.84 (SD=5.431) and the mean score for the post training Engagement was 25.44 (SD=4.931).

Table 37. Descriptive Statistics for Pre and Post Engagement Scores by Training Condition

	Condition	Mean	Std. Dev.	N	Std. Error	95% CI	
						Lower	Upper
DSSQ Pre-Training Engagement	Virtual	22.84	5.431	32	0.843	21.159	24.528
	Baseline	27.31	3.995	32	0.843	25.628	28.997
	Total	25.08	5.238	64			
DSSQ Post-Training Engagement	Virtual	25.44	4.931	32	0.774	23.891	26.984
	Baseline	29.16	3.743	32	0.774	27.609	30.703
	Total	27.30	4.730	64			

A 2 by 2 between-groups multi-variance analysis was conducted to assess the effect of two training treatments on subject engagement. The independent variables were the baseline training treatment and the virtual treatment. The dependent variables were the pre and post engagement scores from the DSSQ. There was significant interaction ($\alpha = 0.05$) between the training conditions and pre-training engagement, $F(1, 62) = 14.059$, $p = 0.000$ and also for post-training engagement, $F(1, 62) = 11.547$, $p = 0.001$. These results indicate there is interaction of training condition and engagement.

Table 38 provides the descriptive statistics for the DSSQ pre-training worry and post-training worry means and standard deviations for the baseline and virtual treatments, standard error, and 95% confidence intervals. The data, for the baseline group, shows a mean score for pre-training Worry was 11.72 (SD=4.887) and the mean score for the post-training Worry was 10.44 (SD=5.418). The data for the soldiers who received the virtual treatment shows a mean score for pre-training Worry was 13.28 (SD=5.721) and the mean score for the post training Worry was 9.16 (SD=5.023).

Table 38. Descriptive Statistics for Pre and Post Worry Scores by Training Condition

	Condition	Mean	Std. Dev.	N	Std. Error	95% CI	
						Lower	Upper
DSSQ Pre-Training Engagement	Virtual	13.28	5.721	32	0.941	11.401	15.161
	Baseline	11.72	4.887	32	0.941	9.839	13.599
	Total	12.50	5.336	64			
DSSQ Post-Training Engagement	Virtual	9.16	5.023	32	0.924	7.310	11.002
	Baseline	10.44	5.418	32	0.924	8.591	12.284
	Total	9.80	5.223	64			

A 2 by 2 between -groups multi-variance analysis was conducted to assess the effect of two training treatments on subject worry. The independent variables were the baseline training treatment and the virtual treatment. The dependent variables were the pre and post worry scores from the DSSQ. There was no significant interaction ($\alpha = 0.05$) between the training conditions and pre-training worry, $F(1, 62) = 1.380$, $p = 0.245$ nor for post-training engagement, $F(1, 62) = 0.962$, $p = 0.330$. These results indicate there is no interaction of training condition and worry.

Figure 12 provides a consolidated summary of means for the DSSQ Distress, Engagement, and Worry scores for the novice soldier population.

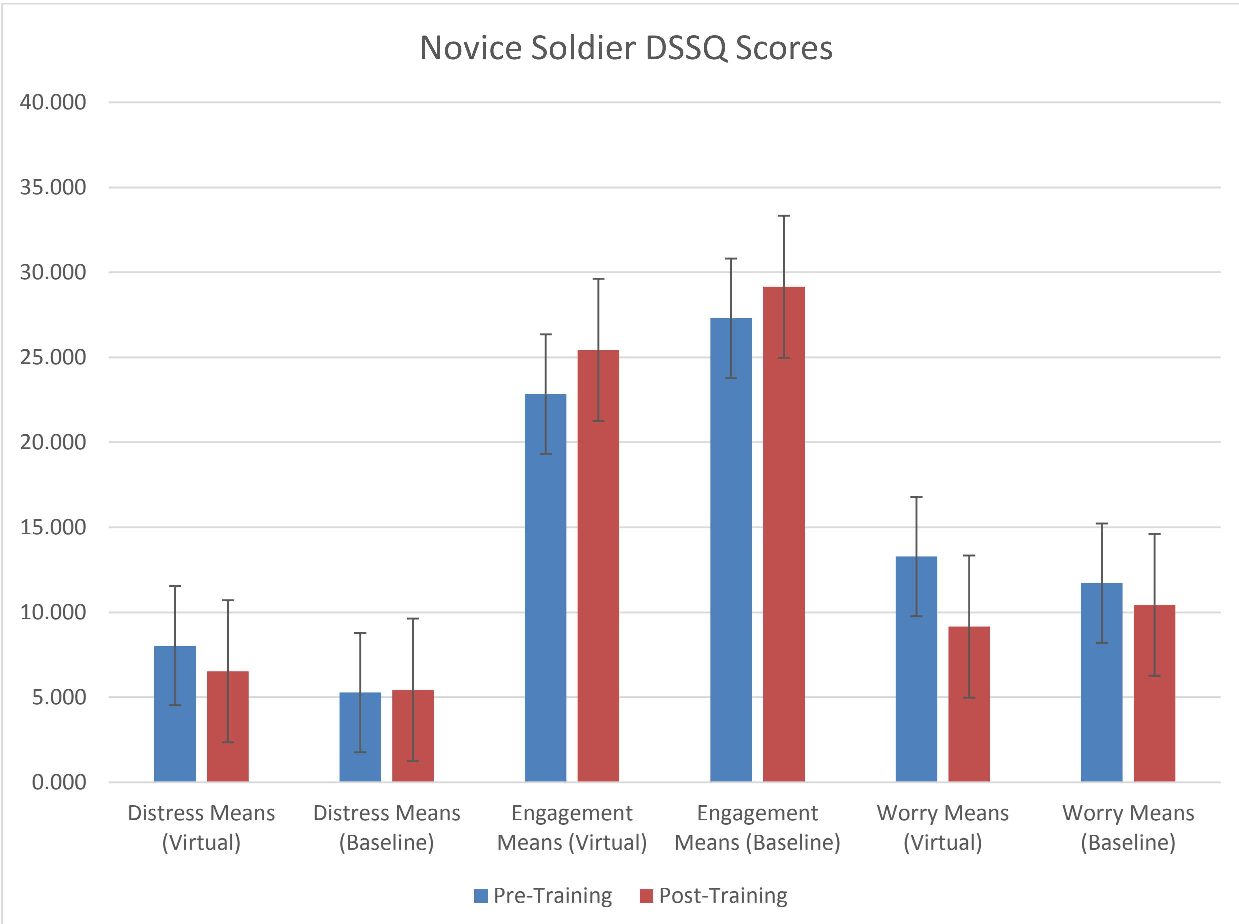

Figure 12. Summary of DSSQ Scores for Novice Population

CHAPTER FIVE: DISCUSSION AND FUTURE WORK

Chapter Five Summary

The purpose of this study was to examine the performance of soldiers trained via two training treatments. The two treatments were a baseline of classroom training using slides and a virtual environment. The soldier population was also divided into novice and experienced categories. This chapter reviews the study conclusions, discusses the lessons learned and future work.

Conclusions

Hypothesis 1: Experienced Individual Performance Effects of Training Treatments

Hypothesis 1 was, "*for the task of clearing a room, there is no difference between performances of experienced soldiers who have been trained using traditional classroom means versus experienced soldiers who have been trained using game based virtual environments*". This hypothesis was tested using a chi-square analysis and a logistic regression. The testing was performed on both the raw data and data adjusted for the removal of outlier participants.

The soldiers were provided two opportunities to run through the live room clearing exercise for performance evaluation. For their first try, the Chi-square analysis of the individual performances resulted in a probability of 0.07, which was higher than our threshold of $\alpha = 0.05$. Although the result mathematically indicates the two conditions are independent and the hypothesis is technically supported, the result is also very close. The logistic regression showed an odds ratio of 39%, meaning the soldiers trained with the virtual means were less likely to obtain a "Go" rating in the live assessment after the first try.

According to both civilian and military experts, the tasks of building and room clearing is a perishable skill (Davis, 2007; Heite, 2010). The key to maintaining and honing those skills is to regularly practice and train. Although the chi-square and logistic regression support the hypothesis, the experienced soldiers who received classroom training numerically did better after the first try. Since 94% of the soldiers in this study had prior training, an explanation for this outcome could be that experience weighed heavily on the performance.

After the second try, all of the soldiers who were provided the virtual treatment passed their live assessment. Only 27 of the 32 soldiers provided the baseline assessment passed their live assessment. The chi-square analysis resulted in a very low probability of 0.0067, which is significant and far below the alpha of 0.05. The hypothesis is now not supported and the indication is that the soldier performance of virtual training treatment was different than the baseline treatment.

The soldiers who received the virtual treatment improved dramatically and as a result all received a "Go" rating on the second try. A possible explanation for this is that the virtual treatment alone was about as effective as the baseline treatment, however the virtual treatment coupled with the feedback of a live instructor may be more effective.

The soldiers who received baseline treatment improved slightly, but still had five who did not pass the second try. An examination of the data revealed that of the five who did not pass, two had no prior room clearing training and had 1 year or less in service. Three of the five had no deployments and had the rank of Private First Class. It should also be noted that all five did not receive a "Go" rating on either attempt. This lack of prior training strengthens the argument that experience influenced the outcome of the performance.

The results of this examination revealed the hypothesis was no longer supported for round 1 assessments and was supported for round 2 assessments after the outliers were removed. This could indicate that the virtual training in absence of guidance of a live instructor may be less effective than traditional classroom methods. Further, the soldiers provided with the virtual treatment and the feedback after the round 1 assessment showed dramatic improvement in round 2. Table 39 provides a summary of the chi-square tests, their associated probabilities and results for both the Round 1 and Round 2 data and the outlier tests for Hypothesis 1.

Table 39. Summary of χ^2 Tests, Probabilities, and Results for Hypothesis 1

	χ^2	p	H_o Supported
Round 1 Data	3.22	0.07	Y
Round 2 Data	7.3557	0.0067	N
Round 1 Data (outlier Removed)	5.25	0.022	N
Round 2 Data (outlier Removed)	3.61	0.057	Y

Statistical Power Analysis for Hypothesis 1

A post-hoc statistical power analysis was conducted for hypothesis 1. The analysis involved utilizing four variables: the sample size of 64 soldiers, one degree of freedom, $\alpha = 0.05$, and medium effect size of 0.3, per Cohen's convention (Cohen, 1992). For this analysis, the statistical package G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) was used. The power value for 64 soldiers is 0.67 with a critical chi-square value of 3.84. Figure 10 shows the central and non-central distributions for the post hoc goodness of fit test. The round one chi-square value was 3.22, which was less than the critical value of 3.84 and the round two chi-square value was 7.35.

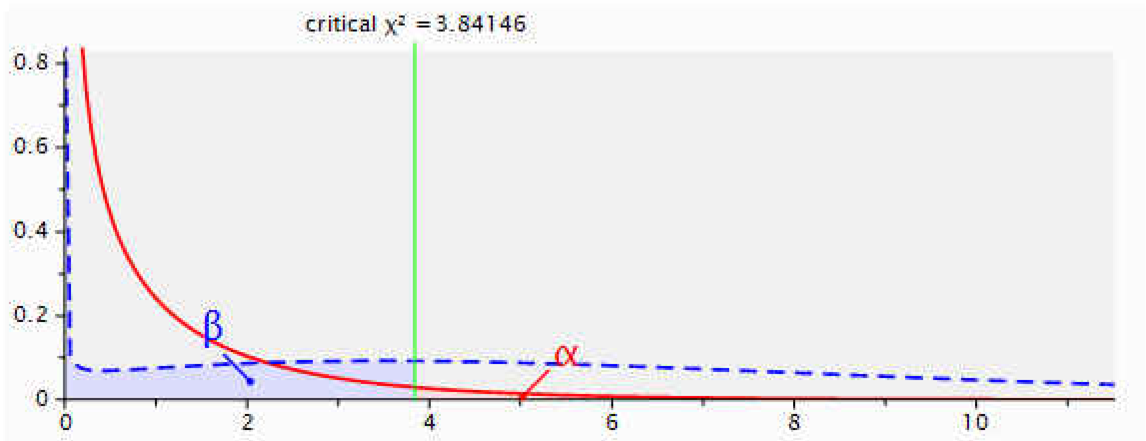

Figure 13. Central and Non-central Distributions for Post hoc χ^2 Goodness of Fit Test

A post-hoc statistical power analysis was conducted for hypothesis 1 with the removal of the two outlier soldiers. The power value for 60 soldiers is 0.64 with a critical chi-square value of 3.84. The round one chi-square value was 5.25, which was more than the critical value of 3.84 and the round two chi-square value was 3.61.

Hypothesis 2: Experienced Collective Performance Effects of Training Treatments

Hypothesis 2 was "*for the task of clearing a room, there is no difference between the performances of teams of experienced soldiers who have been trained using traditional means versus teams of experienced soldiers who have been trained using game based virtual environments*". The subject matter expert provided ratings on 7 of the 8 teams from each treatment (14 total) and only provided one collective rating per team after the second attempt. Video recordings of the live assessments were also made available to the subject matter expert for review to assist in the rankings.

The Mann-Whitney test is based on relative rankings from the subject matter expert in an ordinal fashion. Since there were 16 teams, they were ranked by performance in a scale from 1 to 16 where the lower the ranking, the higher the relative performance. For example a team with a relative rank of 1 did very well relative to a team ranked near the bottom at 13. Although there were only 14 teams evaluated, the ranks were still performed 1 to 16.

The results of the Mann-Whitney test revealed an indication that there was no difference between the performances of the teams based on training treatment. With a calculated probability of 44%, this is much higher than α of 0.05.

A possible explanation for this outcome is that the virtual environment provided adequate interaction capabilities to support team rehearsals of the room clearing task. As part of the virtual training, the soldiers were given five opportunities to practice the task in the simulator. This is different than the normal activity of practicing the task in a taped area in a parking lot or randomly available empty room. The indication here is that the practice tasks performed within the virtual environment provided a similar experience.

Statistical Power Analysis for Hypothesis 2

A post-hoc statistical power analysis for the Mann-Whitney t-test was conducted for hypothesis 2 using the G*Power application. For this analysis, the team's rankings were used in this t-test. The mean ranking for the teams who received the baseline treatment was 8.43 with a standard deviation of 3.77. The mean rankings for the teams who received the virtual treatment was 6.57 with a standard deviation of 4.6. The effect size for the power analysis was calculated to be 0.44 using a two tail test and a normal parent distribution. The power value for the sample sizes of 7 for the two group was calculated to be 0.11.

Hypothesis 3: Novice Individual Performance Effects of Training Treatments

Hypothesis 3 was "*for the task of clearing a room, there is no difference between the performances of novice soldiers who have been trained using traditional classroom means versus novice soldiers who have been trained using game based virtual environments*". This hypothesis was tested using a chi-square analysis and a logistic regression. The testing was performed on both the raw data and data adjusted for the removal of outlier participants.

The cadets were provided two opportunities to run through the live room clearing exercise for performance evaluation. For their first try, 15 of the 32 cadets who received baseline treatment got a "Go" rating and 17 of the 31 cadets who received virtual treatment received a "Go" rating. The chi-square analysis of the individual performances resulted in a probability of 0.3996, which was higher than our threshold of $\alpha = 0.05$. The result indicates the two conditions are independent and the hypothesis is supported. The logistic regression showed an odds ratio of 137%, meaning the soldiers trained with the virtual means were more likely to obtain a "Go" rating in the live assessment after the first try.

For the second try, the chi-square analysis of the cadet's individual performances resulted in a probability of 0.946, which was also higher than our threshold of 0.05. Again, the result is not significant and the hypothesis is supported.

The demographics data indicated that four of the participants had previous room clearing experience. The performance data for these participants were excluded and an additional analysis was performed. In the adjusted data, the chi-square analysis for round 1 performance provides a probability of 0.427. The result is not significant and the hypothesis is still supported. Chi-square analysis for round 2 performance provides a probability of 0.605 and indicates the hypothesis is supported. Table 40 provides a summary of all tests performed for hypothesis 3, including the chi-square values and probabilities.

Table 40. Summary of χ^2 Tests, Probabilities, and Results for Hypothesis 3

	χ^2	p	H_o Supported
Round 1 Data	0.527	0.3996	Y
Round 2 Data	0.0045	0.946	Y
Round 1 Data (outlier Removed)	0.6305	0.427	Y
Round 2 Data (outlier Removed)	0.267	0.605	Y

Statistical Power Analysis for Hypothesis 3

A post-hoc statistical power analysis was conducted for hypothesis 1. The analysis involved utilizing four variables: the sample size of 63 soldiers, one degree of freedom, $\alpha = 0.05$, and medium effect size of 0.3. The power value for 63 soldiers is 0.663 with a critical chi-square value of 3.84. Figure 15 shows the central and non-central distributions for the post hoc goodness of fit test. The round one chi-square value was 0.527, which was less than the critical value of 3.84 and the round two chi-square value was 0.0045.

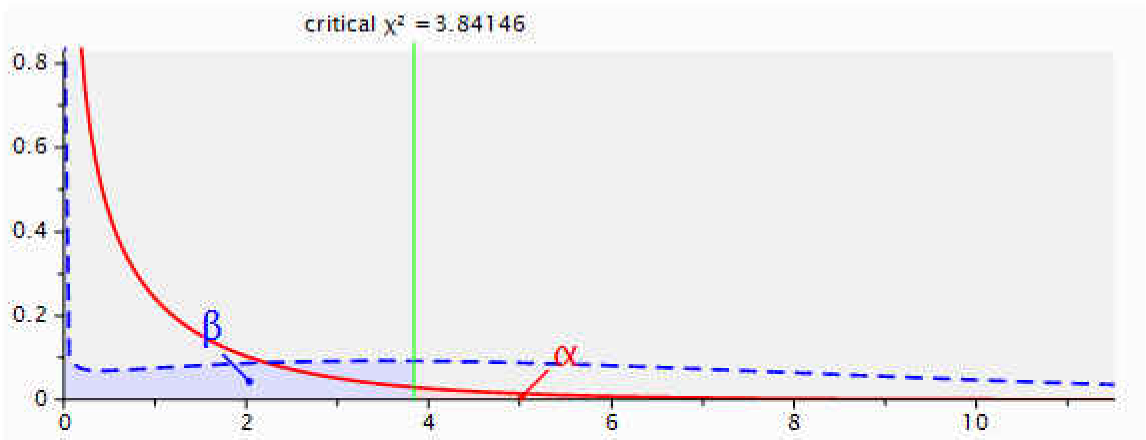

Figure 14. Central and Non-central Distributions for Post hoc χ^2 Goodness of Fit Test

Hypothesis 4: Novice Collective Performance Effects of Training Treatments

Hypothesis 4 was "*for the task of clearing a room, there is no difference between the performances of teams of novice soldiers who have been trained using traditional means versus teams of novice soldiers who have been trained using game based virtual environments.*" The subject matter expert provided ratings on 7 of the 8 teams from each treatment (14 total) and only provided one collective rating per team after the second attempt. Video recordings of the live assessments were also made available to the subject matter expert for review to assist in the rankings.

The Mann-Whitney test is based on relative rankings from the subject matter expert in an ordinal fashion, from 1 to 16 in this case. Although there were only 14 teams evaluated, the ranks were still performed 1 to 16. The results of the Mann-Whitney test revealed an indication that there was no difference between the performances of the teams based on training treatment. With a calculated probability of 37%, this is much higher than α of 0.05.

Similarly with the experienced group, this supports a possible explanation that the virtual environment provided adequate interaction capabilities to support team rehearsals of the room clearing task. As part of the virtual training, the soldiers were given five opportunities to practice the task in the simulator. This is different than the normal activity of practicing the task in a taped area in a parking lot or randomly available empty room. There is also a free play component here that allowed the teams to practice as much as they wanted in the time allotted. The indication here is that the practice tasks performed within the virtual environment provided a similar experience.

Statistical Power Analysis for Hypothesis 4

A post-hoc statistical power analysis for the Mann-Whitney t-test was conducted for hypothesis 4 using the G*Power application. For this analysis, the team's rankings were used in this t-test. The mean ranking for the teams who received the baseline treatment was 7.4 with a standard deviation of 4.65. The mean rankings for the teams who received the virtual treatment was 9.6 with a standard deviation of 4.89. The effect size for the power analysis was calculated to be 0.47 using a two tail test and a normal parent distribution. The power value for the sample sizes of 7 for the two group was calculated to be 0.12.

Hypothesis 5: Experienced and Novice Population Performance Comparisons

Hypothesis five states that "*for the task of clearing a room, there is no difference between the performances of experienced soldiers and novice soldiers who have been trained using traditional means and game based virtual environments.*" To test this hypothesis a series of six chi-square analysis were performed on the combined individual performance data and an additional Mann Whitney analysis was performed on the combined collective data. The first two chi-square tests used performance data from all 127 participants and compared baseline versus virtual treatments for each round of assessment. The next two analysis focused on the 64 novice and experienced soldiers who received the baseline treatment, the performance data from round 1 and round 2 assessment was compared. The last two individual soldier analyses examined the performance data for the 63 participants who received the virtual treatment. Lastly, the performance data from 30 teams was analyzed for dependence.

All 127 of the soldier's performances in round 1 assessments was pooled and analyzed to compare the baseline to the virtual treatments. With a chi-square probability of 0.427, there is a strong indication that the treatments made no difference in the outcome. The hypothesis is supported for the first round assessment. Similarly, the round 2 results were pooled and analyzed to compare baseline to virtual treatments. The chi-square probability goes down a bit to 0.227, but this is still safely above $\alpha = 0.05$ indicating the treatments are still independent and the hypothesis is supported. This particular test could be useful for decision makers to determine if the virtual technologies could be safely included as part of a future curriculum. Further, since the indication is there is no difference in performance between either experience levels, the treatment could be included at any time in the training cycle for use as a pre-training treatment or as a skills maintenance tool.

The next two individual performance analyses focus on the soldiers who received the baseline treatment only and looked for differences in performance based solely on experience differences. The round 1 assessment results of the two participant pools yielded a chi-square probability of 0.042, just below α threshold of 0.05. This indicates the hypothesis is not supported and the experienced group did significantly better than the novice group. Interestingly, the analysis of the round 2 results have a chi-square probability of 0.522, well above $\alpha = 0.05$ and the conclusion again supports the hypothesis.

The round 1 performance data from 63 soldiers who received the virtual treatment was analyzed to determine if there would be differences based on experience. The chi-square analysis indicates a probability of 0.7, well above the threshold of 0.05 indicating the hypothesis is supported and there is no difference in performance between the novices and experienced soldiers. After round 2, the chi-square probability analysis produced a very small result of 0.004. This is a strong indicator the hypothesis is not supported and is backed up by the data which shows the experienced soldiers in this group all passed the live assessment after round 2.

Table 41 provides a summary of the individual performance tests performed, the number of participants in each test, the chi-square value and probability and the test result.

Table 41. Individual Performance Hypothesis Test Summary

	n	χ^2	*p*	H_o Supported
All Soldier Performance: Baseline vs Virtual Treatment Round 1	127	0.63	0.427	Y
All Soldier Performance: Baseline vs Virtual Treatment Round 2	127	1.456	0.227	Y
Baseline Treatment: Experienced vs Novice Round 1	64	4.146	0.042	N
Baseline Treatment: Experienced vs Novice Round 2	64	0.4103	0.522	Y
Virtual Treatment: Experienced vs Novice Round 1	63	0.1478	0.7	Y
Virtual Treatment: Experienced vs Novice Round 2	63	8.129	0.004	N

Statistical Power Analysis for Hypothesis 5

A post-hoc statistical power analysis for the total soldier performance chi-square analysis performed for hypothesis 5. The analysis involved utilizing four variables: the sample size of 127 soldiers, one degree of freedom, $\alpha = 0.05$, and medium effect size of 0.3. The power value for 127 soldiers is 0.922 with a critical chi-square value of 3.84. Figure 16 shows the central and non-central distributions for the post hoc goodness of fit test. The round one chi-square value was 0.63, which was less than the critical value of 3.84 and the round two chi-square value was 1.456.

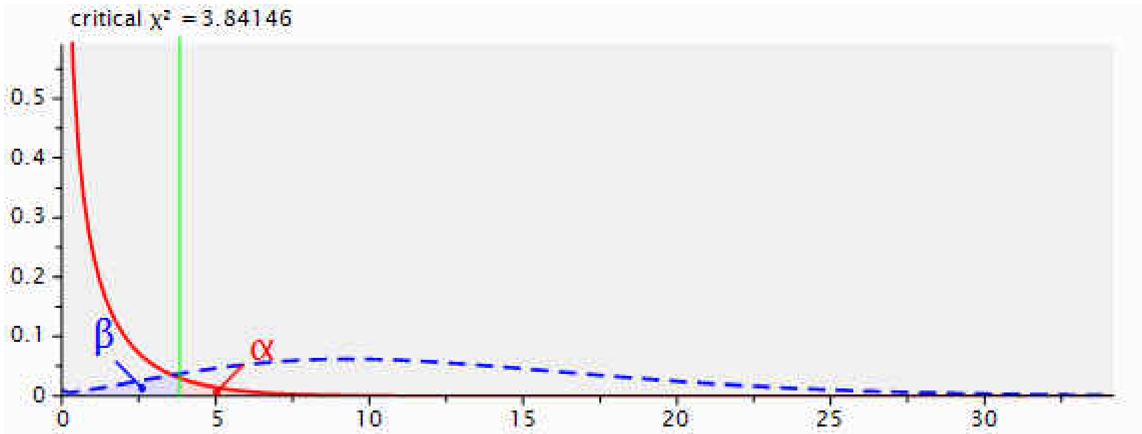

Figure 15. Central and Non-central Distributions for Post hoc χ2 Goodness of Fit Test

DSSQ Analysis

Collective Population DSSQ Discussion

The Dundee Stress State Questionnaire (DSSQ) is a self-reporting mechanism for the participants to report pre-training and post-training stress. This study focused on three of the factors provided by the DSSQ: Distress, Engagement, and Worry. Table 42 provides a summary of multi-variance analysis of variables between subjects (live treatment and baseline treatment) and within subjects (pre-training and post-training).

Table 42. Summary of DSSQ Results, α = 0.05

Population			***F*(1,62)**	***p***	**Training Condition Interaction**
Experienced	Distress	Pre-Training	1.29	0.292	N
		Post-Training	2.676	0.107	N
	Engagement	Pre-Training	0.60	0.442	N
		Post-Training	3.654	0.060	N
	Worry	Pre-Training	0.445	0.507	N
		Post-Training	0.977	0.327	N
Novice	Distress	Pre-Training	4.621	0.035	Y
		Post-Training	0.701	0.406	N
	Engagement	Pre-Training	14.059	0.000	Y
		Post-Training	11.547	0.001	Y
	Worry	Pre-Training	1.380	0.245	N
		Post-Training	0.962	0.330	N

The data collected from the DSSQ could provide some insights into the performance of the soldiers in the live evaluation. In the experienced population, the soldiers who received the virtual training had an overall lower mean of distress than the soldiers who received classroom (baseline) training. Further, the mean distress score for the virtual group declined from pre-training to post-training. The mean distress scores remained almost unchanged for the baseline group. Similarly the worry means went down from pre-training to post-training for the virtual group, vice almost unchanged for the baseline group. The most obvious difference between the virtual and baseline groups were with engagement. Engagement increased for the virtual group from pre-training to post-training condition. This is interesting as engagement decreased for the classroom group. This could be an indication that the virtual group were receptive to the interactive nature of the simulator, versus the passive nature of the lecture before the allotted practice time.

The analysis of the DSSQ data collected from the novice population had significant results, specifically in distress and engagement. The UCF ROTC students indicated higher distress before the virtual training period and lower stress after. The students who received the baseline training had lower distress means. The virtual group's distress means were much higher than the experienced soldier's virtual group's distress means, indicating there may have been increased anxiety about the task they had never been exposed to previously. Mathematically, this translates into a main effect for distress.

The UCF ROTC student's mean engagement scores were counter to that of the Leesburg soldier's. The mean engagement scores for the virtual group were lower than that of the baseline classroom group. The mean scores increased from pre-training to post-training for both treatment groups, also a difference than the experienced soldiers. An explanation for the increase in engagement pre-training to post-training in the virtual group is that the ROTC students may have enjoyed the interactive nature of the simulator. Additionally, the ROTC students had a different instructor than the Leesburg (experienced) group which may contribute to the elevated engagement means for the baseline treatment. These scores resulted in an analysis which shows a significant main effect for distress for the virtual.

Lastly, the UCF ROTC students reported elevated worry means in both training treatments however the worry means also went down in the post-training surveys. This could also be an indicator of anxiety about the task and alleviate of that anxiety after training.

Lessons Learned

The first two studies were conducted using training material derived from the field manual and provided to the 2/124th as part of their normal monthly drill schedule and to the UCF ROTC as an addendum to a regularly scheduled laboratory instructional period. In the future, more studies will need to be performed with minimal impact on the current courses, with minimal interference. Although the power for the individual performance evaluations was quite high (>92%), the power for the collective performance evaluations was low. This is expected as the fire teams had four members each.

The data that was collected used the metrics and performance rubrics traditionally used by the training officers. The use of a "go/no-go" performance metric limited the data analysis to simply determining dependence of the variables up on each other. In subsequent studies, the inclusion of an additional rubric that has a multi-point Likert scale will allow for the determination of "goodness" and perhaps eventually to a return on investment. A measurement of how much of a difference in performance between treatments will allow for the further determination of a return on investment.

There were only slight differences indicated in the performance of experienced soldiers trained using the traditional versus virtual treatments. The logical conclusion can be made that soldiers could be instructed to perform basic room clearing preparation as individuals or as a collective in a virtual environment before reporting to a training center. This virtual preparation could be used to truncate the amount of time needed to complete the onsite instruction or be used to provide additional advanced instruction during the originally allotted time.

Not only does the U.S. military need to understand which training tasks are most appropriate for use in a virtual world, but also when in the training cycle the virtual training belongs. These studies specifically separated the novice soldiers from the experienced soldiers in an attempt to determine if there were any differences in performance based on skill. The rationale was that if a significant difference in performance could be detected between groups, then this could be used as an indicator of where in the training cycle the virtual training could best be utilized.

Referring to the descriptive statistics for the Leesburg 2/124th FLARNG (Figure 6.), an interpretation could be that the virtual condition gave more of a meaningful practice experience. There was a dramatic increase in performance between round one evaluation and round two evaluation for the teams who received the virtual training treatment. This could indicate that coupling the virtual instruction with a live instructor's feedback could be responsible for the increased performance gains. This observation is supported by findings by researchers in the medical community who are using virtual reality technology for surgical simulators (Haque & Srinivasan, 2006). When the outliers removed, this observation is further strengthened by the evaluations of the baseline group, who had even less of a difference in improvement.

The descriptive statistics for the UCF ROTC cadets indicate overall the virtual teams performed slightly better, however the analytical statistics showed no meaningful difference. This could indicate exposure to the virtual early in the training cycle is just as beneficial as using the simulators later for maintenance of skills.

A particularly interesting trend that was observed was in the performances of the experienced and novice teams. Although there were only 14 teams evaluated for the experienced group and 16 teams for the novice group, the Mann-Whitney tests showed that there no difference in performance between the baseline and virtual treatments. This is an indication that the virtual training could be performed in a distributed manner, having teams working and training together in the same virtual environment while located at home station or field element could have a tremendous savings in travel expenses.

The implication is that pre-training tasks could be performed before deployment to the training center, thus saving time and money in the process. Useful distributed training activities could include exposure to the high level tasks in the up-coming courses, acclimation to tactical concepts through interactive role play. For example, a soldier training for a promotion to E5 (SGT) would need to learn the tactics for how to operate as a squad leader in a Warrior Leader Course. A class of E4's could take turns in the simulation based trainer acting as squad leaders while practicing skills by working through simulations in the trainer. The participants in this training do not need to be collocated. The results from this research indicate this kind of distributed training could be done with similar effectiveness as live, when performed with skilled instruction. This new kind of distributed simulation based training approach with skilled instruction could produce soldiers who are better prepared for classroom and live training. This could provide time savings in reduced onsite training, or allow trainers to use the extra time with the introduction of advanced concepts, or simply allow for extra training time in the live evaluation environment.

Perhaps one of the most important lessons learned from this training effectiveness evaluation is how to execute a study this magnitude. The two data collection events required four Institutional Review Board (IRB) determinations from UCF and ARL. It is hoped the requirements for the IRB can be modified in the future so that reciprocity between the institutions can be recognized so that administrative burden may be reduced. In the end, the data collected for this dissertation required the cooperation of Army Research Laboratory staff, University of Central Florida Institute for Simulation Technology's Applied Cognition & Training in Immersive Virtual Environments (ACTIVE) Laboratory, the Leesburg 2/124th Apache Company commander, staff, and soldiers, and UCF ROTC commander, instructors, and cadets. Since the timeline for these studies spanned over a year, it was important to keep regular contact with commanders and team leaders. They were briefed regularly and provided progress updates.

Preparation of the simulator was critical. For both data collection events, the team was allowed to fold into an existing training activity which meant we had a window of opportunity to have access to the soldiers and cadets. It was very important to have a stable simulator with a lean interface. Time to train the soldiers and cadets on the usage of the simulator was reduced by only providing exactly what buttons and interface options that was needed to do the tasks. This allowed the team to train the subjects on simulator as efficiently as possible and get the soldiers and cadets into the task training as fast as possible. For both subject pools this time was reduced to less than 15 minutes each.

Future Work

In late spring of 2015, the team began running training effectiveness evaluations with the 211th Regiment at Camp Blanding's Regional Training Institute. The Warrior Leader Course was selected as the training activity for examination. It was clear from the work performed for this dissertation that more collective team training events needed to be evaluated to gain a clearer picture of the possible effectiveness of the virtual technologies for simulation based training. The future work includes getting access to and evaluating as many teams as possible. Further, the evaluation activities are taking place at higher echelons and examining squad level performance, rather than fire teams.

The effectiveness evaluations will continue in the future with squad and platoon level activities and higher. Eventually, the plan is to have enough data to support reliable conclusions on where in the training cycle the simulation based training technology should be used for infantry soldier skills, which echelons of command are the technologies most effective, and which tasks should be trained with the technology and which should not.

APPENDIX A: PARTICIPANT CONSENT FORM

Informed Consent to Participate in Research Form
Army Research Laboratory
Human Research & Engineering Directorate
Aberdeen Proving Ground, MD 21005

Project Title: Understanding Virtual World Training Effectiveness

Project Number: ARL- W91CRB08D0015

Principal Investigators: Mr. Douglas Maxwell
Human Research and Engineering Directorate
Army Research Laboratory
Simulation Training and Technology Center
(407) 242-5097, douglas.b.maxwell@us.army.mil

Dr. Stephanie Lackey
University of Central Florida
Institute for Simulation and Training
(407) 882-2427, slackey@ist.ucf.edu,

You are being asked to participate in a simulation-based training investigation for Virtual World (VW) training environments. This consent form explains the evaluation and your part in it. Please read this form carefully before you decide to take part. You can take as much time as you need. Please ask any questions at any time about anything you do not understand. You are a volunteer. If you begin this study, you can still change your mind later.

Purpose of the 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 entering and clearing a room.

Test Procedures

You will be asked to complete a demographics questionnaire concerning your military background and experience. You will also be asked to complete a spatial ability test and color vision test. You will view training content for a room clearing mission presented by an instructor or within a VW. Following the brief, you will complete up to 5 trials of a room clearing training scenario in a real world training environment or within a virtual world. You will complete a series of surveys related to your perceived level of usefulness, workload, presence, and engagement during the training scenarios. After the training scenarios, you will receive a post-training brief and complete up to 3 room clearing evaluation scenarios within a real-world setting.

Discomforts and 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).

Benefits

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

Duration

Your participation in this experiment will take approximately two hours.

Confidentiality

Your participation in this research is confidential. The data will be stored and secured in the offices of the principal investigator in a locked file 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. However, complete confidentiality cannot be guaranteed because officials of the U.S. Army Human Research Protections Office and the Army Research Laboratory's Institutional Review Board are permitted by law to inspect the records obtained in this study to insure compliance with laws and regulations covering experiments using human subjects.

Consent to record video, audio and/or photographic data

We would like your permission to take video and audio recordings and/or photographs of your experimental session. Video and audio recordings and/or photographs will be considered privileged and held in confidence. 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 evaluation? *(Check one and initial)*
☐ Yes ☐ No Initial _____

Contact Information for Additional Questions
You have the right to obtain answers to any questions you might have about this research at any time during this test. Please contact anyone listed at the top of the first page of this consent form for more information about this study. You may also contact the Chairperson of the Army Research Lab Institutional Review Board, at (410) 278-5992 or (DSN) 298-5992 with questions, complaints, or concerns about this research, or if you feel this study has harmed you. The Chairperson can also answer questions about your rights as a research participant. You may also call this number if you cannot reach the research team or wish to talk to someone else.

Voluntary Participation
Your decision to be in this evaluation 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. Military personnel cannot be punished under the Uniform Code of Military Justice for choosing not to take part in or withdrawing from this study, and cannot receive administrative sanctions for choosing not to participate. Civilian employees or contractors cannot receive administrative sanctions for choosing not to participate in or withdrawing from this study. You must be 18 years of age or older to take part in this research study. If you agree to take part in this research study based on the information outlined above, please sign your name and indicate the date below. You will be given a copy of this consent form for your records.

This consent form is approved from 11 January 2014 to 12 January 2014.

______________________________________ ____________________
Participant's Signature Date

____
Participant's Printed Name

______________________________________ ____________________
Signature of Person Obtaining Consent Date

____
Printed Name of Person Obtaining Consent

APPENDIX B1: 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 Exempt Human Research

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

To: **Stephanie Jane Lackey**

Date: **January 08, 2014**

Dear Researcher:

On 1/8/2014, the IRB approved the following minor modification to human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Modification Type:	The following research assistants have been added to the study: Crystal Maraj, James Tyson, Sushunova Martinez, Heidi Buehrer, Andrew Watson, Samuel Cosgrove, and Karla Badillo-Urquiola
Project Title:	Understanding Virtual World Training Effectiveness
Investigator:	Stephanie Jane Lackey
IRB Number:	SBE-13-09784
Funding Agency:	Army Research Laboratory(ARL)
Grant Title:	
Research ID:	1052585

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

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

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

Signature applied by Joanne Muratori on 01/08/2014 08:01:06 AM EST

Joanne Muratori

IRB Coordinator

Page 1 of 1

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-PIs: **Douglas B. Maxwell**

Date: **October 23, 2014**

Dear Researcher:

On 10/23/2014, the IRB approved the following human participant research until 10/22/2015 inclusive:

Type of Review:	UCF Initial Review Submission Form Expedited Review Category #7
Project Title:	Understanding Virtual World Training Effectiveness – Phase 2
Investigator:	Stephanie Jane Lackey
IRB Number:	SBE-14-10707
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 10/22/2015, 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:

Signature applied by Joanne Muratori on 10/23/2014 03:25:31 PM EDT

IRB Coordinator

APPENDIX B2: ARL IRB APPROVAL LETTER

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-E 31 October 2013

MEMORANDUM FOR:	Douglas Maxwell, Simulation Training and Technology Center, ARL-HRED, Orlando, FL
FROM:	Paul Rose, IRB Chairperson, Army Research Laboratory
SUBJECT:	Research Exemption Determination

PROJECT TITLE:	Understanding Virtual World Training Effectiveness
PROJECT NUMBER:	ARL 14-002
SUBMISSION TYPE:	Initial Protocol
REVIEW TYPE:	Administrative

The purpose of this memorandum is to notify you that the study identified above has been determined to be exempt from the Common Rule. The purpose of your study is to determine the effectiveness of a virtual environment (VE) training system which is used to train Solders in room clearing tasks. As many as 100 subjects will be recruited and randomly assigned to the VE training system or to classroom training conditions. After training, subjects will engage in room clearing tasks. Questionnaires and tests will also be administered to subjects:

- Demographics Questionnaire
- Cube Comparison Test
- Dundee Stress State Questionnaire (pre- and post-task)
- NASA Task Load Index
- Presence Questionnaire
- Engagement Measure
- Flow State Scale-Short
- Technology Acceptance Measure
- Simulator Sickness Questionnaire

The Common Rule exempts research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods. Your study constitutes a comparison among instructional techniques. In addition, the Common Rule exempts research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless:

(i) Information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and

(ii) Any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

Your study will employ educational tests and will use survey procedures to give insight to the instructional techniques. I have determined that disclosure of the subjects' responses outside the research could not reasonably place them at risk of criminal or civil liability or be damaging to their financial standing, employability, or reputation.

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 changes in the procedures so that a determination may be made that the changes do not alter the determination that the study is exempt research.

I have included a copy of the consent form containing header information with this memorandum. Please use it in your study.

Good luck with your study.

ROSE.PAUL.N.1231792275

Digitally signed by ROSE.PAUL.N.1231792275
DN: c=US, o=U.S. Government, ou=DoD, ou=PKI, ou=USA, cn=ROSE.PAUL.N.1231792275
Date: 2013.11.04 13:51:56 -05'00'

Paul N. Rose, Ph.D.
Chairperson, Institutional Review Board

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-E 8 December 2014

MEMORANDUM FOR:	Douglas Maxwell, MSME, Simulation & Training Technology Center, HRED, ARL, Orlando, FL
FROM:	Theresa M. Straut, CIP, RAC, Human Protection Administrator, Army Research Laboratory IRB
SUBJECT:	Approval of Research, ARL 14-086
PROJECT TITLE:	Understanding Virtual World Training Effectiveness – Phase 2
SUBMISSION TYPE:	Initial Protocol
REVIEW TYPE:	Expedited Review
APPROVAL PERIOD:	08 December 2014 to 07 December 2015

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 08 December 2014.

The following project documents were reviewed on November 18, 2014.

- Protocol 14-086, Understanding Virtual World Training Effectiveness – Phase 2
- Informed Consent form 141023_JS
- DSSQ Short Form
- Flow State Short Scale
- Simulator Sickness Questionnaire
- Cube Comparison Test
- Engagement Measure
- ROTC Demographics Questionnaire
- NASA-Task Load Index
- Presence Questionnaire
- Technology Acceptance Questionnaire
- Study application to University of Central Florida IRB

The expedited reviewer requested changes. The following amended documents were reviewed on December 4, 2014.

- Protocol 14-086, Understanding Virtual World Training Effectiveness – Phase 2
- Cube Comparison Test

The requested CITI certifications for the study team were reviewed December 8, 2014.

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 chairperson:

- 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 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: 2014.12.08 16:11:21 -05'00'

Theresa M. Straut, CIP, RAC
Human Protection Administrator

APPENDIX C: DEMOGRAPHIC SURVEY

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 ____________

What is your MOS? ________________

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

Location	Time
________________________________	_____________________
________________________________	_____________________
________________________________	_____________________
________________________________	_____________________
________________________________	_____________________
________________________________	_____________________

Do you have training experience in room clearing? ___Yes ___No
Do you have training experience in room clearing that used simulation or virtual reality? ___Yes ___No

If yes, what type and purpose?

Type	Purpose
_________________	____
_________________	____

Do you have training experience in building clearing? ___Yes ___No
Do you have training experience in building clearing that used simulation or virtual reality? ___Yes ___No

If yes, what type and purpose?

Type	Purpose
_________________	____
_________________	____

3. Educational Data
What is your highest level of education received? Select one.
____ GED
____ High School
____ Some College
____ Bachelors Degree
____ M.S/M.A
____ Ph.D or other doctorate
____ Other: ______________________

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

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

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 D: DUNDEE STRESS STATE QUESTIONNAIRE

KEY

DSSQ-3 STATE QUESTIONNAIRE

PRE-TASK QUESTIONNAIRE

Instructions. This questionnaire is concerned with your feelings and thoughts at the moment. Please answer **every** question, even if you find it difficult. Answer, as honestly as you can, what is true of **you**. Please do not choose a reply just because it seems like the 'right thing to say'. Your answers will be kept entirely confidential. Also, be sure to answer according to how you feel **AT THE MOMENT**. Don't just put down how you usually feel. You should try and work quite quickly: there is no need to think very hard about the answers. The first answer you think of is usually the best.

For each statement, circle an answer from 0 to 4, so as to indicate how accurately it describes your feelings **AT THE MOMENT**.

Definitely false = 0, Somewhat false = 1,
Neither true nor false = 2, Somewhat true = 3, Definitely true = 4

1. I feel concerned about the impression I am making.	0	1	2	3	4
2. I feel relaxed.	0	1	2	3	4
3. The content of the task will be dull.	0	1	2	3	4
4. I am thinking about how other people might judge my performance.	0	1	2	3	4
5. I am determined to succeed on the task.	0	1	2	3	4
6. I feel tense.	0	1	2	3	4
7. I am worried about what other people think of me.	0	1	2	3	4
8. I am thinking about how I would feel if I were told how I performed	0	1	2	3	4
9. Generally, I feel in control of things.	0	1	2	3	4
10. I am reflecting about myself.	0	1	2	3	4
11. My attention will be directed towards the task.	0	1	2	3	4
12. I am thinking deeply about myself.	0	1	2	3	4
13. I feel energetic.	0	1	2	3	4
14. I am thinking about things that happened to me in the past	0	1	2	3	4
15. I am thinking about how other people might perform on this task	0	1	2	3	4
16. I am thinking about something that happened earlier today.	0	1	2	3	4
17. I expect that the task will be too difficult for me.	0	1	2	3	4
18. I will find it hard to keep my concentration on the task.	0	1	2	3	4
19. I am thinking about personal concerns and interests.	0	1	2	3	4
20. I feel confident about my performance.	0	1	2	3	4
21. I am examining my motives.	0	1	2	3	4
22. I can handle any difficulties I may encounter	0	1	2	3	4
23. I am thinking about how I have dealt with similar tasks in the past	0	1	2	3	4
24. I am reflecting on my reasons for doing the task	0	1	2	3	4
25. I am motivated to try hard at the task.	0	1	2	3	4
26. I am thinking about things important to me.	0	1	2	3	4
27. I feel uneasy.	0	1	2	3	4
28. I feel tired.	0	1	2	3	4
29. I feel that I cannot deal with the situation effectively.	0	1	2	3	4
30. I feel bored.	0	1	2	3	4

KEY

DSSQ-3 STATE QUESTIONNAIRE

POST-TASK QUESTIONNAIRE

Instructions. This questionnaire is concerned with your feelings and thoughts while you were performing the task. Please answer **every** question, even if you find it difficult. Answer, as honestly as you can, what is true of **you**. Please do not choose a reply just because it seems like the 'right thing to say'. Your answers will be kept entirely confidential. Also, be sure to answer according to how you felt WHILE PERFORMING THE TASK. Don't just put down how you usually feel. You should try and work quite quickly: there is no need to think very hard about the answers. The first answer you think of is usually the best.

For each statement, circle an answer from 0 to 4, so as to indicate how accurately it describes your feelings WHILE PERFORMING THE TASK.

Definitely false = 0, Somewhat false = 1,
Neither true nor false = 2, Somewhat true = 3, Definitely true = 4

1. I felt concerned about the impression I am making.	0	1	2	3	4
2. I felt relaxed.	0	1	2	3	4
3. The content of the task was dull.	0	1	2	3	4
4. I thought about how other people might judge my performance	0	1	2	3	4
5. I was determined to succeed on the task.	0	1	2	3	4
6. I felt tense.	0	1	2	3	4
7. I was worried about what other people think of me.	0	1	2	3	4
8. I thought about how I would felt if I were told how I performed	0	1	2	3	4
9. Generally, I felt in control of things.	0	1	2	3	4
10. I reflected about myself.	0	1	2	3	4
11. My attention was directed towards the task.	0	1	2	3	4
12. I thought deeply about myself.	0	1	2	3	4
13. I felt energetic.	0	1	2	3	4
14. I thought about things that happened to me in the past	0	1	2	3	4
15. I thought about how other people might perform on this task	0	1	2	3	4
16. I thought about something that happened earlier today.	0	1	2	3	4
17. I found the task was too difficult for me.	0	1	2	3	4
18. I found it hard to keep my concentration on the task.	0	1	2	3	4
19. I thought about personal concerns and interests.	0	1	2	3	4
20. I felt confident about my performance.	0	1	2	3	4
21. I examined my motives.	0	1	2	3	4
22. I felt like I could handle any difficulties I encountered	0	1	2	3	4
23. I thought about how I have dealt with similar tasks in the past	0	1	2	3	4
24. I reflected on my reasons for doing the task	0	1	2	3	4
25. I was motivated to try hard at the task.	0	1	2	3	4
26. I thought about things important to me.	0	1	2	3	4
27. I felt uneasy.	0	1	2	3	4
28. I felt tired.	0	1	2	3	4
29. I felt that I could not deal with the situation effectively.	0	1	2	3	4
30. I felt bored.	0	1	2	3	4

KEY

Distress = q6 + q17 + q27 + q29 - q2 - q9 - q20 - q22 + 16
Engagement = q5 + q11 + q13 + q25 - q3 - q18 - q28 - q30 + 16
Worry = q1 + q7 + q10 + q12 + q16 + q19 + q21 + q26

Range of Scores Scores will range from 0-32
Remaining Items are candidate items for an imporved worry scale.

REFERENCES

Alexander, A. L., Brunyé, T., Sidman, J., & Weil, S. A. (2005). *From Gaming to Training : A Review of Studies on Fidelity , Immersion , Presence , and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games*.

Ang, C. S., & Zaphiris, P. (2008). Social learning in MMOG: an activity theoretical perspective. *Interactive Technology and Smart Education*, *5*(2), 84–102. doi:10.1108/17415650810880754

Argote, L., & Ingram, P. (2000). Knowledge Transfer: A Basis for Competitive Advantage in Firms. *Organizational Behavior and Human Decision Processes*, *82*(1), 150–169. doi:10.1006/obhd.2000.2893

Avildsen, J. G. (1984). *The Karate Kid*. U.S.A.: Columbia Pictures Corporation.

Beal, S. (2009). EXPLORING THE USE OF A MULTIPLAYER GAME TO TRAIN INFANTRY COMPANY COMMANDERS EXPLORING THE USE OF A MULTIPLAYER GAME TO TRAIN INFANTRY COMPANY COMMANDERS. In *Interservice/Industry Training Simulation and Education Conference (I/ITSEC)* (pp. 1–12). Orlando.

Bezerra, C. E. B., Comba, J. L. D., & Geyer, C. F. R. (2012). Adaptive load-balancing for MMOG servers using KD-trees. *Computers in Entertainment*, *10*(3), 1–16. doi:10.1145/2381876.2381881

Bowers, C. a., Serge, S., Blair, L., Cannon-Bowers, J., Joyce, R., & Boshnack, J. (2013). The Effectiveness of Narrative Pre-Experiences for Creating Context in Military Training. *Simulation & Gaming*, *44*(4), 514–522. doi:10.1177/1046878113475341

Brown, K. G. (2001). Using Computers To Deliver Training: Which Employees Learn and Why? *Personnel Psychology*, *54*(2), 271–296. doi:10.1111/j.1744-6570.2001.tb00093.x

Brusso, R. C., & Orvis, K. a. (2013). The impeding role of initial unrealistic goal-setting on videogame-based training performance: Identifying underpinning processes and a solution. *Computers in Human Behavior*, *29*(4), 1686–1694. doi:10.1016/j.chb.2013.01.006

Burgos, D., Tattersall, C., & Koper, R. (2007). Re-purposing existing generic games and simulations for e-learning. *Computers in Human Behavior*, *23*(6), 2656–2667. doi:10.1016/j.chb.2006.08.002

Cannon-Bowers, J., & Bowers, C. (2009). Synthetic Learning Environments-On developing a Science of Simulation Games, and Virtual Worlds for Training. In S. W. J. Kozlowski & E. Salas (Eds.), *Learning, Training, and Development in Organizations* (pp. 229–261).

Casey, J. (2014). Open Simulator. Retrieved from http://opensimulator.org/wiki/Main_Page

Charsky, D. (2010). From Edutainment to Serious Games: A Change in the Use of Game Characteristics. *Games and Culture*, *5*(2), 177–198. doi:10.1177/1555412009354727

Childress, M. D., & Braswell, R. (2006). Using Massively Multiplayer Online Role-Playing Games for Online Learning. *Distance Education*, *27*(2), 187–196. doi:10.1080/01587910600789522

Cohen, J. (1992). Quantitative Methods in Psychology. *Psychological Bulletin*, *112*(1), 155–159. doi:10.1037/0033-2909.112.1.155

Conle, C. (2003). An Anatomy of Narrative Curricula. *Educational Researcher*, *32*(3), 3–15. doi:10.3102/0013189X032003003

Connolly, T. M., Boyle, E. a., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, *59*(2), 661–686. doi:10.1016/j.compedu.2012.03.004

Davis, K. (2007). Learning from SWAT. Retrieved March 25, 2015, from http://www.officer.com/article/10250052/learning-from-swat

De Freitas, S., & Griffiths, M. (2007). Online gaming as an educational tool in learning and training. *British Journal of Educational Technology*, *38*(3), 535–537. doi:10.1111/j.1467-8535.2007.00720.x

Dev, P., Youngblood, P., Heinrichs, W. L., & Kusumoto, L. (2007). Virtual worlds and team training. *Anesthesiology Clinics*, *25*(2), 321–36. doi:10.1016/j.anclin.2007.03.001

Dickey, M. (2005). Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. *Educational Technology Research and Development*, *53*(2), 67–83. Retrieved from http://link.springer.com/article/10.1007/BF02504866

Dwyer, T., Griffith, T., & Maxwell, D. (2011). Rapid Simulation Development Using a Game Engine-Enhanced Dynamic Geo-Social Environment (EDGE). *The Interservice/Industry Training, Simulation and Education Conference*, (11298), 1–10. Retrieved from http://ntsa.metapress.com/index/M1253340G425U323.pdf

EPIC Games, I. (2014a). Unreal Engine. Retrieved from https://www.unrealengine.com/products

EPIC Games, I. (2014b). Unreal Engine.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. doi:10.3758/BRM.41.4.1149

Fletcher, J. D. (2009). Education and Training Technology in the Military. *Science*, *323*(January), 72–75. Retrieved from www.sciencemag.org

FM 3-21.8 The Infantry Rifle Platoon and Squad. (2007) (Vol. 8). Washington D.C. Retrieved from http://armypubs.army.mil/doctrine/DR_pubs/dr_a/pdf/fm3_21x8.pdf

Fong, G. (2006). Adapting COTS games for military experimentation. *Simulation & Gaming*, *37*(4), 452–465. doi:10.1177/1046878106291670

Gabrielova, E., & Lopes, C. V. (2014). Impact of Event Filtering on OpenSimulator Server Performance. In *Summer Simulation Multi-Conference*. The Society for Modeling and Simulation International. Retrieved from http://www.ics.uci.edu/~lopes/documents/summersim14/gabrielova_lopes_preprint.pdf

Hale, K. S., Stanney, K. M., & Malone, L. (2009). Enhancing virtual environment spatial awareness training and transfer through tactile and vestibular cues. *Ergonomics*, *52*(2), 187–203. doi:10.1080/00140130802376000

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. doi:10.1109/TITB.2005.855529

Heite, R. (2010). Silent but deadly; Marines enhance room clearing skills. Retrieved from http://www.mcipac.marines.mil/NewsCenter/NewsArticleDisplay/tabid/1144/Article/531195/silent-but-deadly-marines-enhance-room-clearing-skills.aspx

Insinna, V. (2013, December). Army , Marine Corps Look for Better Data on Simulator Effectiveness. *National Devense*, 9–12. Retrieved from http://www.nationaldefensemagazine.org/archive/2013/December/Pages/Army,MarineCorpsLookforBetterDataonSimulatorEffectiveness.aspx

IPKeys. (2014). I-GAME. Retrieved from http://www.ipkeys.com/products/igame/

Johnson, W. (2007). Serious use of a serious game for language learning. *Frontiers in Artificial Intelligence and Applications*, 67–74. Retrieved from http://books.google.com/books?hl=en&lr=&id=GEK93NUHdXYC&oi=fnd&pg=PA67&dq=Serious+Use+of+a+Serious+Game+for+Language+Learning&ots=Rqv9cm2xb5&sig=dtShxcLDWRKTptjmgV9W5Y4ir_c

Kaber, D. B., Riley, J. M., Endsley, M. R., Sheik-Nainar, M., Zhang, T., & Lampton, D. R. (2013). Measuring situation awareness in virtual environment-based training. *Military Psychology*, *25*(4), 330–344. doi:10.1037/h0095998

Kark, R. (2011). Games Managers Play : Play as a Form of. *Academy of Management Leaning and Education*, *10*(3), 507–527.

Keh, H., Wang, K., & Wai, S. (2008). Distance-learning for advanced military education: Using Wargame simulation course as an example. *... of Distance Education ...*, *6*(4), 50–61. Retrieved from http://www.igi-global.com/article/distance-learning-advanced-military-education/1735

Kozlowski, S. W. J., Gully, S. M., Brown, K. G., Salas, E., Smith, E. M., & Nason, E. R. (2001). Effects of Training Goals and Goal Orientation Traits on Multidimensional Training Outcomes and Performance Adaptability. *Organizational Behavior and Human Decision Processes*, *85*(1), 1–31. doi:10.1006/obhd.2000.2930

Lackey, S. J., Salcedo, J. N., Matthews, G., & Maxwell, D. B. (2014). Virtual World Room Clearing : A Study in Training Effectiveness. In *Interservice/Industry Training, Simulation, and Education Conference* (pp. 1–11). Orlando. Retrieved from www.iitsec.org

Lampton, D. R., & Parsons, J. B. (2001). The Fully Immersive Team Training (FITT) Research System : *Presence: Teleoperators & Virtual Environments*, *10*(2), 129–141.

LeRoy Heinrichs, W., Youngblood, P., Harter, P. M., & Dev, P. (2008). Simulation for team training and assessment: case studies of online training with virtual worlds. *World Journal of Surgery*, *32*(2), 161–70. doi:10.1007/s00268-007-9354-2

Lyon, J. (2014). Firestorm Viewer. Retrieved from http://www.firestormviewer.org/

Matthews, G., Emo, A. K., Funke, G., Zeidner, M., Roberts, R. D., Costa, P. T., & Schulze, R. (2006). Emotional intelligence, personality, and task-induced stress. *Journal of Experimental Psychology. Applied*, *12*(2), 96–107. doi:10.1037/1076-898X.12.2.96

Matthews, G., Joyner, L., Gilliand, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire- Towards a state "big three"? *Personality Psychology*, *7*, 335–350.

Matthews, G., Szalma, J., Rose, A., Neubauer, C., & Warm, J. S. (2013). *PROFILING TASK STRESS WITH THE DUNDEE STRESS STATE QUESTIONNAIRE*. (L. Cavalcanti & S. Azevedo, Eds.) (Psyphology). Hauppauge: Nova Science.

Maxwell, D., & Ortiz, E. (2013). Military Open Simulator Enterprise Strategy. Retrieved from http://militarymetaverse.org/

Mcalinden, R., Clevenger, W., & Rey, M. (1998). Using Commercial Game Technology for the Visualization and Control of Constructive Simulations. *Database*, *2*, 225–232.

Mendenhall, W., & Sincich, T. (2007). *Statistics for Engineers and the Sciences* (5th ed.). Upper Saddle RIver: Pearson Prentice Hall.

Morris, C., Hancock, P., & Shirkey, E. (2004). Motivational effects of adding context relevant stress in PC-based game training. *Military Psychology*, *16*(1), 135–147. Retrieved from http://www.psycontent.com/index/C1HW491PLT414575.pdf

Morton, Lucious (Department of the Army, HQ Deputy Chief of Staff, G. (2011). *The U.S. Army Learning Concept for 2015*. Fort Monroe.

Morton, Lucious (Department of the Army, HQ, Colonel, GS, Deputy Chief of Staff, G. (2011). *U.S. Army Training Concept 2012 - 2025*.

Nebolsky, C., Yee, N. K., Petrushin, V. a., & Gershman, a. V. (2003). Using virtual worlds for corporate training. *Proceedings 3rd IEEE International Conference on Advanced Technologies*, 412–413. doi:10.1109/ICALT.2003.1215149

Office of Management and Budget. (2014). Department of Defense Federal Budget Fiscal Year 2012. Retrieved from http://www.whitehouse.gov/omb/factsheet_department_defense

Ortiz, E., & Maxwell, D. (2014). Military Metaverse MOSES. Retrieved from http://militarymetaverse.org/

Orvis, K. a., Horn, D. B., & Belanich, J. (2008). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior*, *24*(5), 2415–2433. doi:10.1016/j.chb.2008.02.016

Orvis, K. a., Horn, D. B., & Belanich, J. (2009). An examination of the role individual differences play in videogame-based training. *Military Psychology*, *21*(4), 461–481. doi:10.1080/08995600903206412

Osborn, K. (2015). Pentagon: Sequestration Would Again Reduce Annual Military Training. Retrieved February 19, 2015, from http://www.military.com/daily-news/2015/02/19/pentagon-sequestration-would-again-reduce-annual-military-train.html

Pennell, R. (2003). *Dismounted Infantry Virtual Environment Training Effectiveness Trial (DIVE TET)*. Farnborough, Hampshire.

Peterson, M. (2010). Massively multiplayer online role-playing games as arenas for second language learning. *Computer Assisted Language Learning*, *23*(5), 429–439. doi:10.1080/09588221.2010.520673

Petroski, H. (1985). *To Engineer is Human: The Role of Failure in Successful Design*. New York, New York, USA: St. Martin's Press/New York.

Pickup, S. (2013). *ARMY AND MARINE Better Performance and Cost Data Needed to More Fully Assess Efforts*. Retrieved from http://www.gao.gov/assets/660/657115.pdf

Reynolds, P. F. J. (2009). The Role of Modeling and Simulation. In J. A. Sokolowski & C. M. Banks (Eds.), *Principles of Modeling and Simulation A Multidisciplinary Approach* (pp. 25–43). Hoboken, New Jersey: John Wiley & Sons, Inc.

Richards, D., & Porte, J. (2009). Developing an agent-based training simulation using game and virtual reality software. *Proceedings of the Sixth Australasian Conference on Interactive Entertainment - IE '09*, 1–9. doi:10.1145/1746050.1746059

Roscoe, S. N., & Williges, B. H. (1980). *MEASUREMENT OF TRANSFER OF TRAINING*. *Aviation Psychology*. Aimes: Iowa State University Press. Retrieved from www.isuppress.edu

Rushmer, R. (2006). Armchair warriors. *Professional Engineering*, *19*(20), 40.

Salas, E., Rosen, M. a., Held, J. D., & Weissmuller, J. J. (2008). Performance Measurement in Simulation-Based Training: A Review and Best Practices. *Simulation & Gaming*, *40*(3), 328–376. doi:10.1177/1046878108326734

Scales, R. (2013). Virtual Immersion Training : Bloodless Battles. *ARMY*, (July 2013), 24–27.

Schneider, E. (2004). Death with a Story. *Human Communication Research*, *30*(3), 361–375. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1468-2958.2004.tb00736.x/full

Schubert, T., & Friedmann, F. (2001). The Experience of Presence : Factor Analytic Insights. *Presence: Teleoperators & Virtual Environments*, *10*(3), 266–281.

Shetty, S. (2010). Empowering students with engineering literacy and problem-solving through interactive virtual reality games. *2010 2nd International IEEE Consumer Electronics Society's Games Innovations Conference*, 1–6. doi:10.1109/ICEGIC.2010.5716890

Sitzmann, T. (2011). a Meta-Analytic Examination of the Instructional Effectiveness of Computer-Based Simulation Games. *Personnel Psychology*, *64*(2), 489–528. doi:10.1111/j.1744-6570.2011.01190.x

Slater, M. (1999). Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence*, *8*(5), 560–565.

Spain, R. D., Priest, H. a., & Murphy, J. S. (2012). Current trends in adaptive training with military applications: An introduction. *Military Psychology*, *24*(2), 87–95. doi:10.1080/08995605.2012.676984

Stanney, K. M., Cohn, J., Milham, L., Hale, K., Darken, R., & Sullivan, J. (2013). Deriving training strategies for spatial knowledge acquisition from behavioral, cognitive, and neural foundations. *Military Psychology*, *25*(3), 191–205. doi:10.1037/h0094962

Sterling, B. (2003). War is Virtual Hell. Retrieved from http://www.wired.com/wired/archive/1.01/virthell_pr.html

Virtual Battle Spaces 3 (VBS3). (2010). Retrieved from https://milgaming.army.mil/VBS3/

Voulgari, I., Komis, V., & Sampson, D. G. (2013). Learning outcomes and processes in massively multiplayer online games: exploring the perceptions of players. *Educational Technology Research and Development*. doi:10.1007/s11423-013-9312-7

Wang, Y., & Braman, J. (2008). Extending the Classroom through Second Life. *Journal of Information Systems Education*, *20*(2), 235–248.

Whitney, S. J., Temby, P., & Stephens, A. (2013). *Evaluating the Effectiveness of Game-Based Training : A Controlled Study with Dismounted Infantry Teams*. Edinburgh.

Witmer, B. G., & Singer, M. (1998). Measuring Presence in Virtual Environments : A Presence. *Presence*, *7*(3), 225–241.

Wray, R. E., Laird, J. E., & Nuxoll, A. (2005). for Urban Combat Training. *AI Magazine*, *26*(3), 82–92.

Zielke, M., Zakhidov, D., Hardee, G., & Kaiser, M. (2014). Game Based Simulation for Philippine Post-Typhoon Stability Operations Training. In *Interservice/Industry Training Simulation and Education Conference (I/ITSEC)*. Orlando.