12-1-2015

The Effectiveness, Usability, and Motivational Characteristics of Using Animated Role- Playing Situational Simulation Programs for Air War College Distance Learning Curriculum

Dennis Michael Armstrong
University of Northern Colorado

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

THE EFFECTIVENESS, USABILITY, AND MOTIVATIONAL CHARACTERISTICS OF USING ANIMATED ROLE-PLAYING SITUATIONAL SIMULATION PROGRAMS FOR AIR WAR COLLEGE DISTANCE LEARNING CURRICULUM

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Dennis Michael Armstrong

College of Education and Behavioral Sciences
School of Educational Research, Leadership and Technology
Educational Technology

December 2015

This Dissertation by: Dennis Michael Armstrong

Entitled: *The Effectiveness, Usability, and Motivational Characteristics of Using Animated Role-Playing Situational Simulation Programs for Air War College Distance Learning Curriculum*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in College of Education and Behavioral Sciences in School of Educational Research, Leadership and Technology, Program of Educational Technology

Accepted by the Doctoral Committee

____

Mia Kim Williams, Ph.D., Research Advisor

____

David Kendrick, Ed.D., Committee Member

____

Christine Kyser, Ed.D., Committee Member

____

Jay Schaffer, Ph.D., Faculty Representative

Date of Dissertation Defense _______________________________

Accepted by the Graduate School

____

Linda L. Black, Ed.D.
Associate Provost
Dean of the Graduate School and International Admissions

ABSTRACT

Armstrong, Dennis M. *The Effectiveness, Usability, and Motivational Characteristics of Using Animated Role-Playing Situational Simulation Programs for Air War College Distance Learning Curriculum.* Published Doctor of Philosophy dissertation, University of Northern Colorado, 2015.

This study examined perceived effectiveness, usability, and motivational characteristics of using animated role playing situational simulations for Air War College distance learning (AWC/DL). The AWC/DL curriculum provides education to senior military officers who are geographically dispersed around the world with varying degrees of internet connectivity, creating specific limitations as to what methods of instruction are viable. An additional challenge is the very high student-to-teacher ratio (620:1). Traditionally, the AWC/DL program relied on text-based readings and tests to teach and evaluate students. Simulations might provide a realistic and valuable augmentation to the curriculum. The key potential advantages of adding simulations are increased motivation and better transfer of learning. However, the key disadvantages are the large expenditure of both time and money to develop simulations. The AWC/DL incorporated their original cultural simulation (OS) into the curriculum in January 2008. A second simulation entitled Visual Expeditionary Skills Training (VEST) was added in 2011 as an alternative to the original simulation. Most research on games and simulations analyze younger groups of students, whereas this research focused on AWC/DL students who are typically in their mid-30s to mid-40s. The geographical diversity, age range of the students, potential benefits from simulations, and high costs for creating simulations all justify research in this area. This study surveyed students who completed either OS or VEST. Students rated perceived effectiveness, usability, and motivation using Likert-scale questions. Motivation questions utilized Keller's (2010) 36-item Instructional Materials Motivation survey. Additionally, completion codes, reflecting choices students made in completing the OS, were analyzed. A total of 1,192 surveys and 2,671 simulation completion codes were analyzed using factor analysis, MANOVAs, stepwise discriminant analysis, and chi squared association analysis. Females generally reported lower levels of video game experience than did their male counterparts. The study found statistical significance between usability and gender as well as between usability and video game experience. Males and experienced video game users seemed to find the simulation more usable. However, the estimated effect size was small (< 2%). The analysis found no evidence of an interaction between gender and video game experience. The study did find significant associations between gender and the choices made during the simulation. Additionally, 1,871 comments from open-ended questions were analyzed and although there were issues with both simulations, students tended to view simulations as good learning tools. The OS required users to load the software on their computer, resulting in numerous technical issues. Furthermore, the structure of the OS led some students to be "caught in an endless loop," resulting in frustration that was specifically cited in 10.1% of surveys. The OS comments provided insight on the importance of how instructions are conveyed, how simulation progress is displayed, and the importance of making controls intuitive or automated. The VEST simulation was internet dependent and received low ratings from students in bandwidth restricted locations. However, those with robust internet connections generally found the simulation engaging and valuable.

Both simulations clarified the challenges of using computer simulation in academically isolated and technologically diverse environments.

ACKNOWLEDGMENTS

I would to thank all the people who helped me along the way. First and foremost my wife, Rose, whose love and patience encouraged my progress. To my daughters for their loving support, which made the journey easier to sustain, and to my parents for instilling in me a desire to learn that has driven me through my entire academic career.

I especially want to thank my committee chairs, Dr. Gall and Dr. Williams, for their support and guidance. I would also like to express my appreciation to Dr. Schaffer for his help and advice and the rest of my committee members, Dr. Bauer, Dr. Lohr, Dr. Kyser, and Dr. Kendrick for their support and valuable feedback. I would also like to thank Dr. Ledlow for her encouragement throughout my dissertation journey.

Finally, I want to express my gratitude to Air University for providing the opportunity to conduct this research and extend my thanks to each of the 1,192 military officer participants without whom this research would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Simulations can be traced back to the 7th century when chess was used to simulate war (Bradley, 2006; Murray, 1913). Over the years since, simulations have been examined as a method of instruction and several attractive advantages as well as a few serious disadvantages have been found. Understanding the advantages and disadvantages is critical to deciding whether to create and utilize simulations for education purposes. Although the military has a long history of using simulations for training purposes, it has a very limited experience in using robust role-playing computer simulations for military education. The distance learning environment under which military education must operate creates challenges that could be met by the use of simulations. However, a more thorough understanding of the effectiveness, usability, and motivational aspects of simulations is required to better understand when it is appropriate to invest the money and time needed to create a simulation instead of relying on more traditional methods of instruction.

Background on Simulations

There is no universally accepted definition for simulations, but basic elements are generally accepted. An educational simulation in its basic form contains a representation of an underlying model of something, normally a real-world activity, with which the user interacts to learn (Rieber & Noah, 2008). The goal is to better understand the underlying model. Like simulations, games do not have clear definition. The basic elements of a game are that it has competition and is rule-based. The goal in a game is to win. However, beyond these basic ideas, there is a great deal of overlap between the two including they require strategies, can include chance and consequences, are designed to be fun, and can be based on reality or fantasy (O'Neil, Wainessa, & Baker, 2005). There are programs that involve elements from both lists and can be classified as simulation games. Role-playing programs can fit into this category (Alessi & Trollip, 2001; Faria, 1998).

The advantages of using simulations for education include the ability to promote active learning, improve motivation, enhance the transfer of learning to real world tasks, provide flexibility, accomplish a task multiple times, create a safe learning environment, offer an alternative to traditional tests or written papers for evaluations, and to test new concepts (Alessi & Trollip, 2001; Annetta, 2008; Fong, 2006; O'Neil et al., 2005). This host of advantages certainly justifies serious consideration for the use of simulations in education. However, there are also disadvantages. While the advent of modern computers has removed some of the practical limitations such as the need for physical space to conduct the simulation and the requirement for additional people behind the scenes to run complex simulations, four important limitations still plague simulations: time, money, acceptance by educators, and improper use. The tremendous amount of time and money required to create a robust computer simulation must be weighed against potential benefits (Allessi & Trollip, 2001). Additional research into the appropriate use of simulations might better inform educators, possibly improving acceptance and aiding in the proper design and implementation of simulations for education.

Pedagogically, the case for simulations is compelling. Simulations involve a series of decision loops where a decision is made followed by its active implementation within the simulation, which provides feedback to the learner who then incorporates the feedback into the next decision (Garris, Ahlers, & Driskell, 2002). This learning cycle can create the disequilibrium Piaget (1952) discussed as the core process of how learning takes place. Simulations are also rich in context, which helps learners understand how to utilize their knowledge for real world tasks (Eck, 2006). Context, coupled with the active learning which is also inherent in simulations, creates experiential learning. Experiential learning can be a more effective way to learn because knowledge is stored and later accessed in a way that is consistent with how it is used in the real world (Laurillard, 2001). Motivation is also a major benefit to simulations. The active learning aspect of simulations is inherently motivating because students enjoy doing something better than passively hearing about it (Allessi & Trollip, 2001). If designers incorporate gaming aspects, they can enhance the intrinsic motivation of the experience (Lepper & Chabey, 1985). When simulations are designed properly, they can gain and maintain the student's attention in an engrossing way that can be highly motivating (Norman, 1993).

Although many studies have examined the use of simulations in education, much still needs to be studied. In a general sense, simulations can be evaluated in terms of their effectiveness, usability, and motivational characteristics. Effectiveness measures the student's ability to transfer learned knowledge to real world situations. Usability measures how easy the simulation is to initially learn and how easy it is to navigate through the various parts of the simulation (Virvou & Katsionis, 2006). Motivation looks at the simulation's ability to gain and maintain the student's attention, provide a sense of relevance, instill confidence in the student's ability to complete the simulation, and generate a sense of satisfaction in the experience (Keller, 2010).

Two factors that could possibly influence these measures are gender and video game experience. Males tend to find games relaxing whereas females tend to avoid games for entertainment and are more skeptical about their value for learning (Bonanno & Kommers, 2008). The subject of skepticism is important because a critical issue in the effectiveness of situational simulations is the ability of the student to suspend his or her disbelief. Simulations are clearly not real, but the ability of the student to engage in the simulation with enthusiasm, rather than skepticism, is important. In addition to the possible gender issue with the suspension of disbelief, Hindle (2002) pointed out three properties of the simulation design that could influence the disbelief: a believable scenario, the simulation occurs in a timeframe that is realistic but overly not drawn out, and the simulation has a reasonable level of sophistication. A reasonable level of sophistication involves creating the appropriate level of fidelity when designing the simulation. Fidelity refers to how realistic the simulator is compared to real life. Ironically, a more realistic or higher fidelity simulation does not always result in more effective simulations. Sometimes a simpler simulation can be more effective (Alessi & Trollip, 2001).

In addition to gender, video game experience might also influence the effectiveness, usability, and motivational measures of simulations. Students who play video games at home have a more positive attitude about the use of games for learning (Bonanno & Kommers, 2008). Their experience and the resulting attitude could influence any or all of the measures. How significant these gender and video game experience factors are in the use of simulations for military education has not yet been explored.

Background on Military Education

This study specifically investigated the use of an animated role-playing situational simulation for use in a professional military education (PME) distance learning curriculum. Professional military education is critical to preparing military officers and enlisted personnel for the new roles and responsibilities they will encounter as they progress throughout their careers (Air University, 2006). Professional military education programs are offered via in-resident and distance learning formats. The traditional approach to PME distance learning curriculum has been to use a correspondence method of sending material to the student to read and then having the student take a multiple choice or short answer test that covers the material. It is a self-paced format and until recently, there were no "virtual" classrooms to connect instructors to the students. The Government Accounting Office (2004) noted the correspondence format resulted in a lower quality of education compared to the in-residence format that used a traditional face-to-face method of instruction.

Beyond the format in which the education takes place, it is worth noting that PME distance learning students also have some distinct differences from traditional college distance learning students: they are predominantly in technical occupations, they are essentially required to complete PME programs regardless of their interest level in the subject, and they are in the same age group as their peers. Air War College (AWC) students are generally in their mid-30s to mid-40s.

Several challenges are involved in PME distance learning that impact the options available to improve the quality of distance learning education. First, there is a much higher student-to-instructor ratio in distance learning compared to the in-residence program. The AWC in-residence program has seminars with approximately 10-15 students and one instructor. The AWC distance learning program has about 5,600 students and approximately nine instructors, which equates to about 620 students for each instructor (J.D. Carlin, personal communication, February 22, 2010). This prevents the use of typical synchronous or asynchronous interaction with instructors. Current manpower issues will likely prevent any change to this ratio. Second, students are literally stationed all over the world, which creates a student body that is essentially spread across every time zone. Finally, the military has recently fought two wars simultaneously and is currently engaged in operations around the globe. This highlights the critical need for a distance learning program with a great deal of flexibility in order for students to be able to work on their PME studies when they get the chance rather than being tied to a specific time to convene online or being held to weekly deadlines for deliverables. The correspondence program has provided this flexibility for students but has not provided the desired level of learning. The incorporation of robust computer-based simulation programs for PME distance learning provides a major opportunity to retain the flexibility PME currently has while potentially achieving a higher level of learning.

The Cultural Simulation Program

Air War College (AWC), the Air Force's senior PME program, implemented a cultural simulation program in January 2008. This robust simulation is differentiated from previous Air Force PME simulations in two ways. First, it was designed to take hours rather than minutes to complete. Second, it was designed specifically to provoke thought by presenting the student with choices that did not have a clear right or wrong answer. Earlier simulations tended to test knowledge or provide a mechanism to change the amount or type of forces to create a successful outcome to a scenario. Air War College students are stationed in many different locations including remote locations with limited computer capability. Previous simulations needed to be more simplistic to match the limited computer capabilities of the students at that time. This new cultural simulation took advantage of improved computer capability of the students and was the first major role-playing, animated computer simulation program for Air Force PME. The program puts the student in the role of a military commander setting up airfield operations in central Africa for a humanitarian mission. The student must interact with local nationals and United Nations personnel from several other countries to work through a number of challenges presented in the scenario. The simulation has the look and feel of an animated role-playing video game. The AWC simulation has not been thoroughly analyzed to evaluate the effectiveness, usability, and motivational characteristics of this approach. That examination is vital to deciding if these types of simulation programs are desirable alternatives to the traditional correspondence approach.

Purpose

The purpose of this study was to examine the effectiveness, usability, and motivational aspects of animated role-playing situational simulation programs for use in professional military education distance learning programs. Motivation was examined in terms of Keller's (2010) four characteristics: attention, relevance, confidence, and satisfaction. Additionally, this study examined what aspects users found valuable or problematic with using the simulation for educational purposes.

Study Rationale

The use of robust animated role-playing situational computer simulations is new to professional military education distance learning programs. Pedagogical theory indicated these simulation programs should enable higher levels of learning to take place, more effective transfer of knowledge and higher motivation, but the theory also acknowledges there has to be an acceptance of the artificial environment or, in other words, an ability to suspend disbelief in order for simulations to be effective. Prior gaming experience and gender might help or hinder effectiveness, usability, and motivation when using role-playing simulation programs for education. Bonanno and Kommers (2008) found evidence that women tend to have less enthusiasm and more skepticism than men about using games for educational purposes. That skepticism could impact their ability to suspend disbelief. Lower enthusiasm and higher skepticism might impact the effectiveness, usability, and motivations aspects of simulations. Likewise, higher video game experience might impact the usability of simulations by making it easier to learn to use and navigate through the simulation. Higher video game experience is also thought to be indicative of a more positive attitude about simulations, which in turn can increase motivation and effectiveness.

Although there are compelling advantages to using simulations, there are disadvantages as well. As Alessi and Trollip (2001) noted, situational simulations are not commonly used for education. They pointed to three factors: the expense involved, the difficulty in creating the programs, and possibly an element of skepticism on the part of educators regarding their effectiveness. They pointed out the use of simulations for education is an area that needs more research (Alessi & Trollip, 2001).

Use of simulations clearly has numerous potential benefits for PME. However, due to the high cost in terms of money and time involved with creating robust simulations for PME distance learning, it is critical that the decision to create these programs be based on a thorough understanding of the benefits. The costs are more obvious and easier for decision-makers to quantify. However, the benefits merit further examination, specifically in the areas of effectiveness, usability, and motivational characteristics of these situational simulations, in order to better understand their potential application for professional military education.

Research Questions

The following research questions guided this study:

Q1 Is the cultural simulation program perceived as effective?

Q2 Does prior gaming experience impact perceived effectiveness, usability, attention, relevance, confidence, and satisfaction when using role playing simulations for PME?

Q3 Does gender impact perceived effectiveness, usability, attention, relevance, confidence, and satisfaction when using role playing simulations for PME?

Q4 Is there an interaction between gender and gaming experience on perceived effectiveness, usability, attention, relevance, confidence, and satisfaction when using role-playing simulations for PME?

Q5 What aspects of the simulation do participants find valuable or problematic?

Definition of Terms

Air War College (AWC). The Air Force's senior professional military education institution. The AWC prepares senior officers and U.S. civilians "for the responsibilities of strategic leadership in joint, interagency, and multinational environments" (Air University, 2006, p. 9).

Effectiveness. A measure of the student's ability to retain and successfully apply newly learned knowledge in a real setting.

Motivation. Reiser and Dempsey (2007) stated, "[m]otivation refers to a person's desire to pursue a goal or perform a task, which is manifested by choice of goals and effort (persistence plus vigor) in pursuing the goal." (p. 84). With regard to educational curriculum, Keller (2010) noted four characteristics of motivation (attention, relevance, confidence, and satisfaction) that are key to understanding and evaluating its application.

Original simulation (OS). The AWC/DL program's original animated, role-playing simulation used to teach cultural understanding.

Professional military education (PME). Prepares "junior, midcareer, and senior noncommissioned and commissioned officers, and selected civilians for progressively more responsible positions throughout the Air Force and DOD" (Air University, 2006, p. 119). The curriculum at each level "builds on the education provided at the previous level" (Air University, 2006, p.119). The five core areas consist of the following: "the profession of arms, military studies, international security studies, communication studies, and leadership and management studies" (Air University, 2006, p. 119).

Usability. For this study, usability is defined as a measure of how easy it is for students to initially learn to use and then subsequently navigate within the program.

Visual expeditionary skills training (VEST). A culture simulation created by the Air Force Culture and Language Center.

CHAPTER II

REVIEW OF LITERATURE

Introduction

Over 40 years ago, Piaget (1951) noted the linkage between play and learning. Modern computers now allow us to use simulations and games to truly take advantage of this link. Although computer simulations have been part of science classrooms since the 1980s, there is still much to learn about the use of simulations and games for educational purposes (Blake & Scanlon, 2007). This examination of the literature begins with a look at what defines games and simulations. There is no consensus to the multiple definitions; in fact, there is some level of overlap between games and simulations. Next, the various advantages and shortfalls of games and simulations are covered followed by an examination of the pedagogical foundation behind the value games and simulations bring to the learning environment. Possible ways to evaluate games and simulations are then examined. Following that, research studies grouped into categories are presented: those dealing primarily with effectiveness, those dealing with motivation, and those dealing with usability. Studies pertaining to the military are covered--looking first at the use of simulations and games and then looking at studies examining professional military education (PME) itself.

Defining Simulations and Games

The challenge in defining games and simulations, and in particular the difference between the two is a lack of consensus on common definitions. A basic definition for a simulation is anything that “simulates” reality. This can help identify activities that are not simulations, such as a teacher lecturing in a traditional classroom. However, this definition is not specific enough to distinguish simulations from activities, like watching a video, which could be a substitute for the real environment. A video lacks interaction--a key ingredient to the value of simulations in an educational setting. Alessi and Trollip (2001) defined an educational simulation as “a model of some phenomenon or activity that users learn about through interaction with the simulation” (p. 213). This definition adds clarity to distinguish simulations from other activities. Interaction is crucial to the educational value of a simulation because it forces the user to make choices and deal with the consequences. That experience is both integral to how simulations work and a key element of the learning process. The student's interaction with the underlying model creates an active learning process, which is further examined later in this chapter.

As noted, there is no one common definition of games. However, as a starting point for this discussion and based on their review of eight different definitions, Salen and Zimmerman (2004) have defined a game as "a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome" (p. 80). The term "artificial conflict" both differentiates games from real life and also indicates an element in a game that has to be overcome through cooperation or competition to achieve an outcome. The defined rules and quantifiable outcome provide the structure and purpose or goal of the game.

A review of the similarities and differences between games and simulations can help provide additional understanding of these definitions. When comparing a simulation to games in a meta-analysis, O'Neil et al. (2005) found a wide variety of definitions for both and many commonalities in definitions of games and simulations. They found in general both games and simulations incorporated actions of the user with actions of another entity, whether computer or human. Both games and simulations have rules, require strategies that can include chance and consequences, are designed to be fun, and can be based on reality or fantasy. They also noted a number of differences between the two. The goal of a game is to win, whereas the main goal of a simulation is to discover cause and effect relationships. Games have competition, either against an opponent or against the game itself. A simulation does not require competition. Games have a linear structure in which the first action influences subsequent actions. Rieber and Noah (2008) provided a simpler description of differences by noting games involve competition and are rule based, whereas simulations put the user into a role and are based on an underlying model that is the foundation of the simulation. Alessi and Trollip (2001) identified many of the same characteristics of games but further noted simulations can be categorized into four different types. Physical simulations present the user with a simulation of an object with which they can interact over time. This allows them to see the effect of those inputs. An example would be simulating a city road system and being able to see the change in traffic patterns as additional roads are constructed. The second type is iterative simulations. They are distinct from physical simulations due to the manner in which the user provides inputs to the simulation. Instead of providing inputs as the simulation unfolds, iterative simulations allow the user to repeat the simulation numerous times using different initial inputs. An example is a simulation of a jet engine in which the user iteratively sets different initial fuel mixtures and runs the simulation to see the impact on engine performance. The third type of simulation is a procedural simulation that focuses on learning the steps needed to complete a task. An example is a flight simulator used to teach basic procedures like starting engines and even complex procedures such as dealing with aircraft control malfunctions in flight. The fourth type is situational simulations that "deal with the behaviors and attitudes of people or organizations in different situations" (Alessi & Trollip, 2001, p. 224). This type is usually a role-playing simulation in which the user takes on a role within the simulation and interacts with other characters.

Although there can be many ways to categorize games, Dempsey, Lucassen, Hynes, and Casey (1996) identified simulations as one of eight types of games they were studying. Like the definitions, the method of categorizing games is also not standardized. It is worth noting that Dempsey et al. found subjects felt simulations needed "clear goals and objectives" (p. 7). This was consistent with Salen and Zimmerman's (2004) definition of a game that requires a quantifiable outcome. Although Salen and Zimmerman did not address simulations as a category of games, they did discuss role-playing games as a category that tested the boundaries of their definition of game in a way similar to Dempsey et al.'s caveat on simulations. Salen and Zimmerman pointed out many role-playing games do not have an endpoint so they would appear not to be games by their definition. However, they concluded even role-playing games with no specific endpoint could fit in their definition of a game because players typically have short-term goals they are trying to achieve. These short term goals might be set by the game or by the players themselves.

In addition to definitions and categories as a way to describe simulations and games, an examination of the process of *gamification* can provide a fuller understanding of game characteristics. Kapp (2012) provided the following definition for gamification: "Gamification is using game-based mechanics, aesthetics, and game thinking to engage people, motivate action, promote learning, and solve problems" (p. 10). This process attempts to apply the beneficial characteristics of games to non-game activities to improve the experience or outcome. Examining the terms of this definition provides more insight into games themselves. "Game-based" refers to having the basic elements of a game, matching those found in the definition of a game: conflict, rules, and outcomes. Kapp noted that "mechanics" refers to elements of game playing such as points, levels, and time limitations. Although these elements do not by themselves make the activity engaging, they are necessary to facilitate that process. "Aesthetics," which emphasizes the appearance of the gamified interface, is important because it impacts users' acceptance of gamification. Kapp noted that "game thinking" is probably "the most important element of gamification" (p. 11). This is the transformation of a non-game activity into a game-like engaging activity. Kim (2011) identified three types of social actions for game thinking: competition, cooperation, and self-expression. In addition to creating the social actions, designers need to maintain the game-like activity over time. Csikszentmihalyi (1997) used the term *flow* to describe the need to balance increasing skills with increasing challenges, thus preventing either overwhelming or boring the user. The final three terms of the definition-- "motivate action, promote learning and solve problems" (Kapp, 2012, p. 10)--refer to the benefits of gamification that match the pedagogical benefits discussed later in this chapter.

Based on these definitions, categories, and characteristics, some games are clearly games (such as Scrabble®) and some simulations are clearly simulations (such as an iterative jet engine simulation). However, even without intentionally applying gamification principles, clearly programs such as those involving role-playing could be categorized as either or both. Recognizing the differences between simulations and games are blurred in some cases, Faria (1998) noted some programs can be identified as “simulation games.” In cases where the program cannot clearly fit into one category, it is reasonable to examine it as both a game and a simulation.

History of Simulations and Games

Before examining the theoretical underpinnings of the educational value in using simulations for educational purposes, it is worthwhile to consider the history of simulations within a learning environment. Using the broadest definition of simulations being anything that simulates reality, the history educational simulations can be traced back to the Greeks’ use of play as a method of instruction (van Ments, 1995). Use of play as a way to teach might not fit many of the definitions of a simulation, but it does illuminate the history of some of the pedagogy associated with simulations. Although the Greek example only loosely fits the definition of a simulation, by the 7th century, chess was used to simulate war (Bradley, 2006; Murray, 1913). Chess does fit within most definitions of a simulation. Similar to actual war, chess players must develop strategies utilizing combinations of various offensive and defensive capabilities to defeat the enemy and might need to sacrifice valuable "pieces" to accomplish that objective. Chess also fits the definition of a game because it involves competition with a winner and loser and is based on clear rules. During the 19th century, Prussia updated the use of simulations for war by incorporating maps and appropriate game pieces for that time era. These "war games" were used for military education of Prussian officers. In the 1950s, the Rand Corporation created simulations for international relations. "In the 1960s, policy planning and research simulations were applied to such fields as health care, transport, welfare, town planning and finance" (Cruickshank & Telfer, 1980, pp. 76-77). Simulations continued to be utilized for education. Simulations used specifically for cross cultural training are particularly germane to this paper's study, which is based on a cultural simulation. Fowler and Pusch (2010) noted that in the United States, culture simulations were used in the 1970s. While there was no universal acceptance of simulations for this purpose, "many had discovered the power of simulation games to give participants an opportunity to experience aspects of encountering and communicating in a culture different from their own" (Fowler & Pusch, 2010, p. 101). Simulations used during the 1950s to 1970s were generally classroom exercises and not computer-based simulations. These non-computer simulation exercises could be very labor intensive activity for an instructor to prepare and execute. Current availability of computers and creation of computer-based simulation programs makes it easier for instructors to utilize this method of learning.

Advantages of Simulations and Games

There are many purported advantages of simulations. A discussion of their pedagogical benefits is covered in the next section, but an overview of the list of those advantages is warranted. Probably the most significant advantages are improved motivation and better transfer of learning (Alessi & Trollip, 2001; Annetta, 2008; O'Neil et al., 2005). Improved motivation stems from factors such as active participation and relevance. Active participation of students making choices within the simulation is thought to be more motivating than passive activities such as reading text or listening to a lecture. Both relevance and transfer of learning are enhanced by the realism and context of the actions taken in the simulation that are similar to actions taken in real life. Using the knowledge within an appropriate scenario helps students understand why the knowledge is beneficial and also helps students practice using the knowledge in the same way they would in real life. Beyond the major benefits of simulations are additional advantages such as (1) improving efficiency, which is measured as more learning per unit of time; (b) adding the flexibility to address learning on several different levels (Allessi & Trollip, 2001); (c) providing the ability to accomplish the event multiple times to test differing strategies; (d) facilitating the pursuit of complex approaches to the learning process; (e) enabling coverage of cognitive and affective domains of learning (O'Neil et al., 2005); and (f) creating the flexibility to address concepts in ways traditional education cannot (de Jong & van Joolingen, 1998).

These advantages deserve closer scrutiny with regard to their pedagogical basis, but it is also important to examine the purported disadvantages of simulations and games. The disadvantages are high monetary cost in creating the simulation, large investment of time required by subject matter experts and programmers to create the simulation, and, perhaps even more critically, the improper use of simulations. In many cases, current use of the simulations and games in education is failing to reach its potential. They are simply used as an added event to an otherwise complete curriculum and typically involve low budget programs that are not interesting or thought provoking (Gonzalez & Blanco, 2008; Lynch & Tunstall, 2008; Skiba, 2008; Wideman et al., 2007). In fact in many cases, these multimedia environments serve simply as a depository to place information for the student to access. Nelson, Bueno, and Huffstutler (1999) referred to this phenomenon as infotainment rather than being a pedagogically-driven design.

Bonk, Kim, and Zeng (2006) stated the importance of simulations and their capability are only now being fully understood. Because the basis for many of the advantages and disadvantages of simulations and games is nested in their pedagogical foundation, it is important to examine that foundation to better understand the proper use and appropriate ways to evaluate simulations.

Pedagogy

A look at the pedagogy of simulations and games needs to start with the context of how simulations and games relate to other methods of instruction. Although it might appear the use of games and simulations for educational purposes is a relatively new phenomenon, it is important to remember that “traditional” education with classrooms and printed books is only about 300 years old. “Before the time of printed books, learning was done primarily through questioning, storytelling, imitation, practice, and play” (Prenski, 2001, p. 91). These older methods were essentially dropped because they were not as practical in the age of mass education as classrooms, lectures, and books. Mass education took on a factory model mindset with students being viewed as a product and standardization of the product being a goal. The assembly line incorporated common curricula such as books and utilized standardized testing as a method of quality control on the "product" being produced by the educational factory (Serafini, 2002). However, Cordova and Lepper (1996) noted educational material utilized for this mass education was decontextualized, which reduced motivation and usefulness. Simulations and games have the capability of addressing this shortfall. Modern computers are now making it possible to merge the need for mass education with the older, more contextualized methods of learning. This has the potential to provide higher motivation levels and greater ability to apply the knowledge gained from an educational lesson to real life tasks.

The Learning Cycle

Piaget (1952) focused on how children learn from play. Although his efforts focused on children specifically, the concepts he developed also apply to how adults learn through simulations and games. He theorized that people learn through a cycle of assimilation--wherein our brains store new ideas in familiar or known categories and accommodation and wherein we have to modify what we know to accommodate the information that does not fit into our known categories. When there is a conflict between assimilation and accommodation, the result is cognitive disequilibrium. Learning takes place as the disequilibrium is resolved. Annetta (2008) pointed out simulations do in fact provide a cycle of assimilation-accommodation conflict to create disequilibrium and, therefore, work to promote learning. Eck (2006) specifically noted that games are an effective learning environment because they are engaging and require frequent use of decision-making. Requirements for frequent decision-making and subsequent feedback are key components to disequilibrium and learning. Garris et al. (2002) created a model to capture this disequilibrium cycle of games. Their model involved a cycle of user “judgments,” creating “behavior” or inputs to the game that resulted in “feedback” from the game that formed the basis for revised "judgments"; the cycle was continued throughout the game playing. This model was fundamentally based on the assimilation-accommodation cycle. As the users received feedback, they need to use the feedback to confirm or adjust their understanding of the situation. Eck (2006) further noted that successful games not only should be specifically designed to create the disequilibrium, they should be specifically designed not to overwhelm the player but keep the user engaged and motivated. Failing to achieve the right level of challenge for users not only decreases their motivation but impacts their learning as well. If it is too easy, they will not be challenged with any significant disequilibrium to resolve and get bored. If it is too difficult, then they will get frustrated with trying to resolve the disequilibrium. Either way, the level of disequilibrium is linked to changes in both motivation and learning.

Context

In addition to the cycle of disequilibrium, context is another pedagogical benefit of simulations. By design, simulations provide context for taking action. When a simulation replicates a real world phenomenon, the context allows students to experience how to apply the new knowledge in the real world. Van Merrienboer and Kirschner's (2007) recommendation for instructional design favored learning tasks in a real or simulated environment because this context helped the student learn the whole task rather than compartmentalizing the knowledge by focusing on learning pieces of the task. By learning to apply the knowledge as they would in the "real" world, students are better able to transfer the knowledge to applications outside of the classroom. Eck (2006) noted learning is more effective when it occurs within context and that relates to its use in the real world. Because context enables knowledge to be learned in the manner in which it was normally used, context results in more effective learning. Van Merrienboer and

Kirschner noted the experiences of accomplishing tasks in a realistic way helped learners create effective schemas and enhanced transfer of knowledge. The environment facilitates the learning process and the ability to apply the new knowledge appropriately in a real world situation. This environment helps decrease the risk of creating inert knowledge in which the learner knows something but cannot apply it in a useful way (Renkl, Mandl, & Gruber, 1996).

An important aspect of context for simulation is the fidelity of the simulation. Alessi and Trollip (2001) identified fidelity as a measure of how closely the simulation reflected the real world. Sedlack (2007) studied the issue of fidelity in a medical simulation and found the limited fidelity of the simulator negatively impacted the transfer of knowledge to real world tasks. Specifically, their simulator test group was less effective in subsequent real world procedures than the control group, which did not receive simulation training. Problems with patient comfort and intubation were negatively impacted by the unrealistic nature of the simulation for both those areas. Clearly a higher fidelity simulator was needed for this training. However, Alessi and Trollip pointed out one of the benefits to simulators was the ability to simplify some tasks that could clearly reduce fidelity and aid in learning difficult tasks, so there was a tradeoff in deciding on the desired level of fidelity to design into the simulation.

Active Learning

Allessi and Trollip (2001) defined the term *active learning* to indicate an environment in which students must take actions (including cognitive actions) rather than passive learning, which simply requires the student to observe. Smaldino, Lowther, and Russell (2008) stated active learning involved participation by the learner and feedback to the learner. Active learning is an inherent property of simulations because simulations require an action (input) by the user and then provide the user with feedback (output) the user then considers before taking additional actions. The number of inputs and outputs vary by simulation and how it is used, but simulations are not a passive activity. Active learning enables students to stay motivated and helps them learn by improving comprehension and recall (Alessi &Trollip, 2001).

Experiential Learning

Butler (2006) noted a link among active learning, the learning cycle, and context. Linking the three together enhances learning as the student creates meaning from situations or context by reconciling the unknown with the known. Simulations take advantage of this link. The combination of context, active learning, and the learning cycle can be referred to as experiential learning, which is another major benefit of simulations. As previously described, simulations can create a situation or context similar to how the student will actually apply the knowledge in the real world. This context can be coupled to the active learning inherent in simulations. Piaget's (1951) disequilibrium that can be cyclically induced by simulators can provide active learning. Laurillard (2001) made the point that simulations can involve experiential learning (active learning within context) and can be an effective way to learn. The primary advantage is learners actively store and later access the knowledge in a manner consistent with how it is used. As a result of experiencing the knowledge within context, it is easier to access the knowledge when confronted with a real life situation that mirrors the experience in the simulation. The design of the simulator with the proper fidelity is critical to creating this synergy of active learning and context.

Motivation

Enhancing motivation is one of the most well-known advantages of simulation-based learning. However, there are many aspects to motivation and, therefore, many approaches to describing the inherent motivational characteristics of simulations. Alessi and Trollip (2001) noted several factors that make simulations motivating for students. Active learning is generally more exciting than passive learning; simulations can change over time and provide more difficult scenarios or problems to maintain a motivating level of challenge for the student. Because simulations provide learning within context, they are seen by learners as more relevant, thus aiding motivation. Gonzolez and Blanco (2008) asserted that a well-designed game can meld education and games seamlessly and results in intrinsic motivation. They went on to identify the term *coherent interface* to describe a program that results in users not being aware they are even interacting with a program. Norman (1993) described this as an "engaged state of focused attention," which can be very motivating, and he explained that this "focused concentration is easiest to sustain when in an experiential mode" (p. 31), which is more common in games but should be applicable to instruction as well. Although it can be challenging to design a simulation that achieves a high level of engaged attention if achieved, the user is likely to perceive it as exceptionally motivating.

Reflection

It is important from a pedagogical perspective to note that while motivation and the combination of active learning and context are important, a successful learning experience also requires a debriefing or reflection mechanism to be effective. Lantis (1998), in his presentation of the development of a role-playing simulation for international relations, pointed out the importance of debriefing as an essential element of the learning process. He noted that a great deal of learning takes place afterward rather than during the event. Reflection or debriefing can help promote and solidify this learning. Cooper (1998) noted that reflection is a critical part of the learning process. Adaption that takes place through the disequilibrium process to a large degree takes place during reflection after the actual experience has occurred. Jong, Shang, Lee, and Lee (2008) similarly noted that reflective learning makes the event even more effective. Unfortunately, as Rieber and Noah (2008) noted, many times reflection is overlooked or omitted from educational games, resulting in less effective learning.

Flexibility

The flexibility to control events in the simulator is another stated benefit (Alessi & Trollip, 2001). Van Merrienboer and Kirschner (2007) noted simulated environments offer more favorable opportunities for learning than real environments because of the control instructors have including the ability to slow down time, repeat the task, and practice rare or costly events. Practicing rare or costly events clearly provides a great benefit to using the simulation. The student can practice applying the knowledge and then experience the ramification of their actions via the simulation rather than in the real world where there may be little likelihood of experiencing the event, the event may be very costly, or the event might result in tragic consequences if handled incorrectly. An example would be pilots using a simulator to practice how to deal with a loss of multiple engines in flight. Although this is a very rare event, pilots need to practice so they are able to safely handle the situation should it occur. The simulator allows the instructor to let the pilot take the scenario to its logical conclusion--landing safely or crashing--based on the pilot's actions. The pilot can repeat the scenario if needed to learn how to correctly handle the emergency. While the flexibility to practice rare events and repeat events provides a clear advantage, the merits of adjusting the passing of time within a simulation are less obvious and can actually have positive or negative impacts on learning depending on how time is used.

Zagal and Mateas (2010) identified four different types of time frames common for video games: real world, gameworld, coordination, and fictive. These define the ways time can be used within a game or simulation; however, a single game or simulation can use more than one of these time frames.

Real-world time matches the way time progresses in the real world--events are tied to the passage of hours or days in the real world. An example would be the massively multiplayer, online role-playing game *Clash of Clans* in which players build defenses and armies, attack other players, and defend against attacks by other players. Each defensive structure has a real-world time period to complete the building. If a player starts a building that takes two days to complete, then regardless of whether the player continues actively playing the game or turns the game off, two real world days elapse before the building is done. Another example of real time in *Clash of Clans* is each time a player starts a raid, he/she has a maximum of three minutes to complete the raid.

Gameworld time is not linked directly to real-world time. The sequence of events establishes passing of time. An example would be *The Sims* in which the player must sequence actions for their Sim character including periodic feeding and rest. If the player makes his/her character repeatedly study or exercise and does not feed the character, then after performing the action several times, the character will refuse to study or exercise anymore and demand to be fed. Another common example would be a card game where the length of the round is not determined by real time but by the speed at which the cards are played. When the round is complete, the next round can begin.

Coordination time deals with multiple players or computer opponents and coordinates when events can take place. An example can also be found within *Clash of Clans*. Clans do not have to participate in clan wars; however, if the clan leader decides to start a clan war, then a 24-hour period of preparation is followed by a 24-hour period of fighting. Each clan can have up to 50 players, all potentially in different countries, making attacks that must occur during the 24-hour time period of real time.

Fictive time uses labels for time passage that are understood by the players but do not match real world time. An example would be *The Sims* where actions take place and a clock moves forward denoting the passage of gametime but the "normal" rate at which the clock progresses is not tied directly to real-world clocks. The player can change the passing of time from a normal rate to fast or even pause the passage of time. When the player logs off, time stops for these Sims and begins again when the player logs on. The type of action taking place in the game or simulation should be reflected in how fictive time is used. For a game that involves building a castle, rounds of action should be labeled as years to denote the length of time it takes to build a castle. "Labeling the rounds in a game as 'days' or 'years' changes a player's expectations of the granularity of action that can be accomplished in a round" (Zagal & Mateas, 2010, p. 850).

Zagal and Mateas (2010) pointed out the time frame used for the game or simulation needs to be appropriate for the technology involved and the type of simulation. Technology might negatively impact experience when using real time in a game or simulation. If the game or simulation uses real time and a delay occurs, either due to the low processor speed of the computer or a slow internet connection, there might be a negative impact on the user's experience, which in extreme cases might make the game or simulation unusable. The opposite could also be true. For example, modern processors are so fast many older computer games tied to processor speed are now unusable without modifications because the action takes place faster than any user can react. Likewise, when multiple players are playing online card games together, players generally must wait for other players to play their cards before taking a subsequent turn. If the delay is excessive, the experience seems less real and less enjoyable. An option could be to place a real-world time limit on the length of time players have to accomplish their turn before a default action takes place, e.g., losing a turn for that round. Changing the rate at which time passes in the simulation or game based on action taking place is another method to better match the technology and simulation. For example, selecting a weapon when the user is attacked might involve several steps on the computer such as opening the bag, viewing the weapons, and selecting the weapon. If the computer did not freeze time, the player might be defeated before he/she could select a weapon. Stopping the passage of time could allow users to accomplish actions that due to technology and the interface might take more time than they would in the real world. Zagal and Mateas used the term *temporal anomalies* to indicate the use of timeframe that creates dissonance with real-world time. The anomalies could cause a negative impact on the user experience with the game or simulation or, as with the example of selecting a weapon, it could be beneficial to the user's experience.

Consideration of the passage of time is important to the design of the game or simulation. As Zagal and Mateas (2010) pointed out, "[t]he very concepts of 'action,' 'event,' and 'influence' require an account of temporality in games--the myriad ways that temporal structure informs gameplay"(p. 845).

Dual Coding

Another pedagogical benefit to simulations is they typically utilize the multimedia effect of dual coding. Dual coding can have a very positive benefit to the effectiveness of learning. Dual coding theory states that using multiple senses during the learning process helps learners create enduring knowledge from the experience (Clark & Paivio, 1991). An example of this is simultaneously receiving inputs both visually and verbally. The information processing of these two inputs is separate and additive. Therefore, consistent or complementary inputs using both forms allow the brain to process the information more effectively. Although other delivery methods can and do use dual coding, it is important to note the multiple forms of inputs received in typical simulations help provide effective learning and make the material more interesting to the learner, thus improving motivation (Clark & Paivio, 1991; Krain & Lantis, 2006). As with many other aspects of simulations, one that is poorly designed can impede learning via dual coding if the two sources of inputs compete for attention from the same processor. For example, if an audio narrative is presented explaining a screen animation, this creates a positive dual coding environment. However, if the designer also provides the text of the audio narrative on the screen, it can result in splitting the visual processor's attention between the text and the animation. This redundant text can create less effective learning (Mayer, Heiser, & Lonn, 2001).

Scaffolding

Another way simulators incorporate educational theory is scaffolding. Van Merrienboer and Kirschner (2007) noted simulators provide a venue for building up from simpler tasks to more complex tasks. They recommended a good way to approach learning tasks is to provide a low fidelity experience and then gradually increase the fidelity as the learner gains experience. Vreman-de-Olde and de Jong (2005) also found simulations could utilize scaffolding for students to deal with problems or subjects that are initially too complex and likely to overwhelm the student. An example of this would be to initially use a flight simulator without radio calls, weather problems, or system malfunctions so the student can learn basic aircraft controls before dealing with those additional elements. Then as the student's experience increases, those elements can be added.

Evaluating Simulations

While there is a strong pedagogical basis behind the use of simulations and games for education, improper use has been noted as one of the disadvantages. Therefore, it is vital to examine how simulations and games can be evaluated to ensure they strongly contribute to the learning process rather than simply serve as "add-ons" with little or no added educational value. The first issue that needs to be examined is whether a comparison study is a valuable approach to evaluating simulations and games. The real issue in a comparison study is whether or not media even matter to learning. This issue is best characterized by the views expressed by Clark (1994) and Kozma (1994) in their respective articles on the subject. Clark contended that media do not affect learning. He uses the analogy of a delivery truck, pointing out the type of truck does not impact the nutritional content of the food inside the truck. His contention is in cases where media have been identified as making a difference, it is really the method of instruction underlying the media that makes it more or less effective. Clark did agree different media could be more efficient in terms of time spent transferring knowledge, and cost could be a factor as well. However, he asserted a lot of time and energy was wasted in researching the differences between the media.

Kozma (1994) disagreed with that contention. While he acknowledged a vast number of studies have indicated media do not matter, he said we should continue to research the issue of differences in media. His position was the learner and the environment are linked in learning. He took a constructivist approach diametrically opposed to Clark's delivery truck approach. He asserted that failing to examine the media risked ignoring a valuable area for research.

These two articles showed the essence of the debate. The issue of comparing computer simulations directly to classroom instruction would necessitate a look at the underlying methods of instruction. Essentially, media cannot be directly compared. The Air War College distance learning (AWC/DL) simulation would have to be compared to the text-based, traditional correspondence program, which is the only other type of media used for AWC/DL. A direct comparison would simply serve as an examination of the underlying method of instruction, which in itself is clearly significantly different. While some comparison of cost or motivation might be appropriate, a direct comparison to effectiveness of instruction would be questionable at best. Instead, effectiveness needs to be measured against the intended learning outcome of the program.

Given that a direct comparison to a different type of instruction is a contentious and possibly ineffectual approach, it is necessary to look at a broad view of how simulations and games can be evaluated without using a comparison approach. O'Neil et al. (2005) used Kirkpatrick's (1996) four levels of evaluation for training and applied them to evaluating games for their study. Kirkpatrick's (1996) four levels of evaluation are reaction, learning, behavior, and results.

Reaction evaluation involves looking at the student's satisfaction with the program or, in other words, their reaction to it. Learning evaluation measures improvements in attitude, knowledge, or skill based on the program. Both of these levels can be evaluated during or immediately after the program. Behavior evaluation examines the change in behavior as a result of accomplishing the program. In other words, it evaluates whether learning can be effectively applied to new situations on the job. Because the behavior evaluation is looking at the transfer of learning, it needs to be evaluated on the job rather than during or immediately after the training. Results evaluation relates to improvements to the organization's bottom line. This involves a much higher level examination and clearly must be accomplished after the fact. Results evaluation looks at cost effectiveness, quality, decreases in accidents, and other performance measures. The higher the level, the higher the cost and effort required to conduct the evaluation, but those higher levels can also yield more valuable results. However, when planning to use behavior or results evaluation, Kirkpatrick (1996) noted lower levels should still be evaluated to provide a better understanding of the results.

Another method to evaluate media was presented by Lohr (2008). She presented three broad categories of effectiveness, appeal, and efficiency to capture the value of a program. Effectiveness examines how well the program works instructionally to impact learners' knowledge. Appeal evaluation deals with the motivational aspect of the program. Do users like it? The efficiency evaluation can be looked at as a test of usability. Is the program intuitive for the learners? Do they spend a great deal of time learning the program rather than the material? While simpler than Kirkpatrick's (1996) levels, it has the advantage of directly assessing the efficiency or usability issue.

Because these broad categories cover the pedagogical issues fully, they are used as a framework for examining current research in the next three sections. However, the names of two of the categories were changed to reflect language commonly recognized within professional military education. Motivation was used instead of appeal and usability was used instead of efficiency. The three categories are effectiveness, motivation, and usability.

Effectiveness

Several aspects of simulation effectiveness for learning have been examined through various studies. Major areas that have been examined are context and associated situated learning, empathy created by role-playing programs, the need for flexibility, and measurements of effectiveness.

Context

Context involves students understanding what they are doing and how it factors into the real world. Research indicated it is an important factor in effectiveness. Barab et al. (2006) examined fourth graders and the impact of context on learning. They noted the value of a simulation is related to the context. A richer context improved learning. They also found the ability of students to generalize their understanding, instead of just being able to apply it to one specific circumstance, was a significant factor in transferability of knowledge. They described inert knowledge as knowledge that could not be applied beyond the particular instance. They found the key to creating generalizable knowledge was the ability to get users to engage and use their knowledge on several different levels. Specifically, they found a need to balance formal structure with abstract principles in creating the framework of the content area. They referred to this as formalism and noted its relationship to context. Too much emphasis on the structure resulted in a school-like program that could decrease motivation and too little structure (or too much emphasis on context) resulted in less efficient learning. Inefficient learning is the result of too much time spent dealing with context. An ideal balance creates relevant formalism that results in an effective transfer of learning (Barab et al., 2006).

Pedaste and Sarapuu's (2006) study, which examined 65 teams of students on virtual hikes through Estonia, identified a need to establish a support system in conjunction with the context to help balance student improvement and help motivation. Their study pointed to the need to have adaptive feedback rather than generic or predefined feedback to improve the effectiveness. Adaptive feedback provided better context for learning by tailoring the feedback to individual student actions.

Finally, many games and simulations have opening sequences normally involving a video that plays when the game or simulation is initially started. The opening sequence can be costly to produce but is normally included to provide additional context to get the user mentally engaged with the game or simulation and understand its purpose. Procci, Lakhmani, Hussain, and Bowers (2014) conducted experiments to examine whether the opening sequence to a game provided context that specifically improved understanding of the goals or improved learning. They did not find any significant difference between their control group, which did not view an opening sequence, and the treatment groups regarding how they performed within the game or what they learned from the game. This could potentially decrease the development cost of games or simulations without impacting learning.

Role-Playing

Krain and Lantis (2006) studied college students accomplishing a simulation exercise to explore the value of role-playing simulations for enhancing their educational experiences. They found the role-playing simulation did enhance the educational experience. It required critical thinking and students gained deeper insight. They also found much of the experiential learning that occurred took place after the exercise rather than during the exercise. This highlighted the importance of reflection as an element to incorporating simulations for learning. Their study used a pre- and posttest format, asking students to evaluate their own knowledge, and found these simulations created active learning, which improved comprehension, problem solving, and retention of material. It is important to note they also found the knowledge gained was about the same for simulations as it was for lectures. This matched the points made by Clark (1994) about media comparisons not yielding any difference in learning. However, interestingly, Krain and Lantis's study found the role-playing simulations resulted in students developing greater empathy, which could potentially make role-playing simulations valuable for certain areas such as learning about culture or dealing with patients. Additionally, they found overall students enjoyed learning via simulations more than traditional methods and could use their new knowledge to address other types of problems.

Bos, Shami, and Naab (2006) conducted a study looking at business students in a role-playing simulation designed to teach ethics. The simulation contained some cultural aspects and did not have any set correct answers. It emphasized perspective-taking as a crucial part of learning ethics. The simulation was set in a foreign country and involved a U.S. company having to deal with various ethical dilemmas resulting from conflicts between U.S. culture and the foreign country's culture. During the simulations, students needed to identify and then explain what actions they would take. The researchers found that having the students explain their choice for each decision improved the learning effectiveness. This activity essentially forced the student to reflect, which has been previously found to be a critical part of the learning process. Similar to Krain and Lantis's (2006) study, Bos et al. also found the use of role-playing in simulations helped develop empathy.

Pacala, Boult, and Hepburn (2006) conducted a study to examine the effectiveness of an aging game in helping medical students better understand the perspective of their elderly patients. They found the use of role-playing helped instill an attitude change in students by creating memorable experiences. Students found the simulations both motivating and worthwhile. Their study concluded the value students found in the simulation was tied to the experiential nature of the simulation and also to post-simulation discussion. Again, the need to couple simulations with post-simulation discussion or reflection was deemed significant.

Flexibility

Blake and Scanlon's (2007) meta- analysis investigated the results of three studies using computer simulations in distance learning for undergraduate science classes. Each study involved a different simulation program. Their analysis identified three features for effective use of simulations in distance learning. First, it was important to be able to tailor the simulation to the student's ability level. If the simulation was too simple, it might actually confuse students by containing elements that did not match well to the real world. If it was too complex, it might also impede their learning. The flexibility to adjust the simulation to provide the appropriate challenge to match a student's ability was critical to simulator effectiveness. Second, multiple representations should be included to allow students to examine the material from different perspectives. The flexibility to provide these different perspectives enhances their understanding of the material. Third, student support is a vital aspect of creating effective learning. Simulations cannot just be handed to students with the expectation that learning will occur. Students need to be given a plan or direction to begin the simulation and the simulation should contain a "help" capability to provide students with information as they need it. Overall, they found simulations motivated students and created a faster transfer of learning.

Measuring Effectiveness

O'Neil et al.'s (2005) meta-analysis study looked at empirical research examining adult learning through games. They examined 19 studies conducted over a 15-year time period. The studies examined used a variety of methods to evaluate effectiveness including performance on the game, observations, surveys, and retention tests. While there was some indication effectiveness of the game and the amount of time and intensity with which learners played the games were linked, it did not necessarily link to more effective learning of skills or knowledge. It might simply have been a measure of the motivational aspect of the game. Consistent with Clark's (1994) contention, O'Neil et al. also found the simulations were not more effective than other methods of instruction.

Douglas, Miller, Kwanza, and Cummings (2007) conducted a study to examine student perceptions of the usefulness of simulators for hospitality business management in higher education. Use of student perceptions was selected over measuring learning more directly because of the challenges involved with direct measurement (control and experimental groups). They used Likert-scale questions on the perceived usefulness of the simulation and found the simulation was perceived overall as useful in helping students develop various skills needed in hospitality management such as planning, decision-making, and understanding hospitality. They also found those students who enjoyed the simulation did not necessarily spend more time using the simulation. Although not directly assessed during the study, they pointed out the possibility that those who liked the simulation might have been more familiar with gaming environments, which might have made them more efficient at learning to interact with the simulation, decreased the time spent, and raised the perception of its value.

Motivation

The idea that simulations can be more motivating for students than traditional classroom environments had its basis in the pedagogical foundations discussed earlier. However, it is important to examine what recent studies have found concerning the motivational aspect of simulations including the influences of experience with video games, gender, and age.

De Leng, Dolmans, Muijtjens, and van der Vleuten's (2006) study on medical students used a virtual learning environment in conjunction with face-to-face classes. The study utilized questionnaires to examine students' perceptions of whether or not the virtual environment stimulated face-to-face discussions and whether it aided student learning. They found evidence that student preference was higher for traditional face-to-face environments rather than online environments. However, they also found multimedia environments were preferred over simple text-based environments. Students felt the multimedia approach was more effective than text in stimulating group discussion. This helped support the concept of the link between dual coding and motivation.

Rieber and Noah (2008) conducted a study to examine adult learning utilizing game activities. The selected game was a simple simulation for students to learn about acceleration verses velocity. The authors found not only did the game produce high motivation, the motivation increased as the students improved at the game. Interestingly, they found evidence that while students enjoyed the simulation, the game seemed to impede explicit learning. However, they also noted that when linked with a visual metaphor that could serve as an organizer for understanding, there was an improvement in tacit learning. While they investigated some of the context of the game, they did not investigate any reflection activity. The game itself did not promote reflective activity. The authors found an outside agent such as a teacher was vital to learning. They concluded that rather than replacing the need for a teacher, use of games increased the teacher's importance to learning since tailored guidance was necessary to ensure understanding. An issue with this finding was the researchers did not include a specific reflection activity as part of the simulation, which might have decreased the importance of the outside agent to the learning process. Finally, they found a need to balance motivation, experiential learning, and reflective learning when designing curricula. All three areas need to be addressed for learning to be effective.

Gender and Video Game Experience

Males and females do not appear to be equal users of video games. Waters (2006) reported on conference comments by David Gardner, Chief Operating Officer for Electronic Arts' worldwide studio, that "40% of teenage girls played video games versus 90% of teenage boys, and most girls lost interest in games within a year" (p. 1). The differences go beyond the rate females play video games compared to males. But as Jenson and de Castell (2010) found looking at 30 years' worth of research on gender and gameplay, many studies did not examine this issue beyond the question of whether females played video games. Specifically, they found many surveys tried to identify whether females played games, but many never examined the types of games chosen, how females played, or the duration of play compared to males.

In a study that did examine differences beyond the rate at which males and females played video games, Greenberg, Sherry, Lachlan, Lucas, and Holmstrom (2010) looked at gender and age in relation to time spent playing video games, types of games the participants preferred, and their motivations for playing those games. They used 1,242 questionnaires completed by 692 public school students in 5th, 8th, and 11th grades and also by 550 undergraduates from two universities. They asked the participants to identity the amount of time they played, their motives for playing, and the types of games they preferred. Overall for the study, males reported spending twice as much time (18.6

hours/week) playing video games as the females (8.2 hours per week). The tendency for males to report spending more time playing video games was consistent across all the age groups studied. The eighth graders reported the most amount of time playing video games. Males also reported stronger motivations for playing games. Females tended to prefer more traditional games such as classic arcade games, cards, trivia, puzzles, and board games. Males tended to prefer physically-oriented video games involving sports, fighting, shooting, and racing. Younger players generally had a fantasy motive and older players generally identified a competitive motive. Analysis explained more variance in game playing for males than females. The authors suggested the lower playing time, motivation, and choice of types of games could be the result of video games being designed from a male perspective. This resulted from issues such as females needing to use male characters to play the games. They noted the male centric design of some video games might explain many of the differences they found. Video game use dropped off above the eighth grade level, but the authors noted this was not a longitudinal study and so the results may have reflected the environment the age cohorts grew up in rather than a trend of decreasing use as people aged.

One specific game that stood out as an extremely popular game played more by females than by males was the long running *The Sims* series produced by Electronic Arts. *The Sims 3* was number four on the top selling computer games of 2012 (Entertainment Software Association, 2013). In Boyes' s (2007) interview with Sharon Knight, Electronics Arts Vice President of Europe Online, she noted 65% of *The Sims* players were female. In a study examining gender identification and nationality, Wirman (2014) looked specifically at *The Sims 2,* the current version of *The Sims* during the time period that study was taking place. *The Sims* was chosen due to the perception of it being a "feminine" game. Wirman's study consisted of only 13 interviews with Finish players who were "game-modifiers," indicating they were devoted players who went beyond just playing the game and actually spent time modifying various items and characters within the game. Eleven of the 13 interviewees had used Finnish-speaking, online communities to help them create game modifications and shared those modifications with the online forum. Despite spending time modifying the games, the females generally did not consider themselves expert gamers. The responses suggested that even though they were devoted players of *The Sims 2,* they did not identify themselves as video game experts, possibly due to the popular image of a devoted player as being a "young male geek." Almost all the interviewees indicated *The Sims 2* was the only video game they played and saw it as different from other video games. Some used terms like virtual dollhouse to describe the game confirming a "feminine" perception of the game. This perception of a separation of *The Sims 2* from other video games and the tendency not to identify as real gamers indicated there could be a difference on how experienced female video game players might identify themselves on surveys dealing with perceived video game experience levels compared to males. Although the study looked at nationality due to the North American cultural references found within the game, this did not appear to have any effect on the players and many even indicated they did not notice the references.

The impact of gender could also influence the perceived learning value and motivational value of the simulation. Bonanno and Kommers's (2008) study specifically examined how gender influenced gaming competence and the learner's attitude toward using games. They discovered males generally found games relaxing but not only did females generally not find games relaxing, they tended to avoid games as a form of entertainment. Females were also generally more skeptical about the value of games for learning. In fact, although females generally viewed games as simply another way to learn, males generally viewed games as more unique and special for learning. Both genders thought games could enhance learning. Males also tended to view games as something to master and generally had the confidence to pursue self-teaching games. Females tended to view games as tools to assist in learning and generally lacked confidence in using self-teaching games. The authors recommended the logon screen include a way for the users to designate whether they were male or female and then design the program to provide a different level of support based on the gender. The authors also found prior game experience was an important factor in attitudes toward the learning game. Students who were enthusiastic about playing video games at home were very positive about the use of games in school. Students who were moderate gamers tended to be positive while students who were non-gamers tended to be neutral or even negative about the use of games for learning.

Although Annetta (2008) focused on age and the use of game playing for learning, he also identified a gender-related issue, noting when girls played games they tended to prefer role-playing games. The bulk of his research focused on the need to use games for learning to connect with and motivate the “Net Generation.” His research found the game playing population is generally 10 to 34-years-old and that a majority of gamers are 14 to 19-years-old. As noted earlier, designers must bear in mind that not all children are experienced computer users; but as the percent of experienced users goes up, the need to create more interactive environments for learning also increases.

Measuring

Gonzalez and Blanco (2008) investigated ways to examine motivation within games. They utilized three general factors of motivation identified by McKeachie (2002) and applied them toward games. First was the expectative factor, which dealt with the players' expectations of their probability of finishing the game. Second was the value factor of the game including intrinsic ("desire for learning") and extrinsic ("reward for learning"). Third was the affective factor, which looked at the "emotional response" to the game (Gonzalez & Blanco, 2008 p. 402). While this model broke out motivation for games into different parts, it did not lend itself to measurement by quantitative means.

Keller and Suzuki (2004) investigated motivation in e-learning using Keller's (2010) attention, relevance, confidence, satisfaction (ARCS) model. The ARCS model identifies four sub-areas of motivation: attention, relevance, confidence, and satisfaction. Attention refers to gaining and maintaining students' attention throughout the lesson. Among the ways in which attention can be gained is the use of interesting visual inputs such as graphics or animation, unresolved problems, and some variability. These can all be easily present in simulations. Relevance can be gained by "authentic" learning experiences that help students see how they would apply the lesson in the real world. This should be inherent within the design of a simulation because simulations are intended to model the real world so students should be able to see how they would actually apply the lesson. Confidence is gained by ensuring students know what is expected of them and feel they can succeed. The instructional designer is key to building confidence rather than the simulation itself but these elements are compatible with simulations. As mentioned, the ability to adjust the level of difficulty within a simulation could help the designer promote student confidence. Finally, satisfaction refers to students having a positive learning experience. Like confidence, the instructional designer is critical to ensuring student satisfaction. The previous three elements could play into students' satisfaction but so could the opportunity to apply what they have learned, recognized, and a sense of fairness (Keller & Suzuki, 2004). Again, these can be consistent with the use of simulations but are contingent on properly designed instruction.

The advantage to Keller's (2010) ARCS model is it can be measured using two established measurement tools--the course interest survey (CIS) and the instructional materials motivation survey (IMMS). The CIS uses a 34-question survey to measure reactions to instruction led by an instructor. The IMMS uses a 36-question survey to measure reactions to instructional material presented without an instructor. These surveys provide established quantifiable measurements that can be utilized for examining motivational characteristics of instruction. Both surveys have established reliability and validity with numerous studies including multiple groups of undergraduate level students as well a validation for use in other cultures.

Usability

Usability looks at the ease or speed by which students are able to use the program to learn. It is important to note usability can indirectly influence both motivation and effectiveness. Usability is influenced by both the actual program and the background of the learners.

The relationship between usability and motivation (likeability) was noted by Virvou and Katsionis (2006) in their study examining the usability and likability of virtual reality games for education of children. The study incorporated multiple data sources including pre- and posttests, interviews, and surveys to evaluate the likeability and usability of the game. In terms of the interaction between the usability and likeability, they found problems with the usability of a game degraded its likeability. Based on the premise that games must be both likeable and usable by a majority of students, they examined the effects of a variety of student backgrounds on usability and likability. Even though the common stereo-type of younger students is they are computer savvy, the researchers found a wide variety of computer skills among children including some who were essentially unfamiliar with computer games.

Several findings Virvou and Katsionis (2006) discovered during their research related directly to usability and likeability. Although experienced computer game users might be perceived as more inclined to find a learning game likeable, that is not always the case. They hypothesized some experienced users might dislike the learning game because they mentally compared it to commercial games they played and found it less satisfying. Overall, familiarity with video games did factor into students' experience in using the games. Generally, novices spent more time learning how to use the games. Novices spent significantly more time incorrectly navigating, resulting in aimless movement or an inability to move. This time was not productive for actual learning of the desired material and also resulted in a higher dropout rate. Interestingly, they found environmental distracters (enticing elements of the game that ware not productive for learning) were more distracting for players with intermediate levels of experience than for either the novices or experts. Overall, the novices found the games more motivating but they actually spent the smallest amount of time (~four hours each) playing the game. They found players at the intermediate level of experience spent more time (~5.5 hours each) playing the games, but they primarily spent the extra time playing with environmental distracters. The experts spent the most time (~6.8 hours each) playing the games. One clear indication of likeability was most of the students indicated they would like to have the game at home to play. The authors found motivation to learn was especially high among students who did not do well in traditional learning environments (poor performers with less discipline). The authors also noted likability seemed to be proportional to the sophistication of the game and having help functions available was critical. Finally, they also pointed out spending more time playing the game resulted in greater exposure to content and usability problems decreased as students spent more time with the game (Virvou & Katsionis, 2006).

Blasi and Alfonso (2006) accomplished a usability study using a prototype simulation called *The Virtual Lab* in high school biology classes. The study looked at the effectiveness of the prototype design, the prior experience of the students, and examined ways to use the simulation to achieve the learning objectives. The simulation had been developed by NASA for use in a classroom environment. Examining the prior experience, they found 64% of the children self-reported they played video games and only 18% played role-playing video games. One key insight was the researchers found some students lacked prior knowledge of basic computer skills the designers expected the students to possess. This clearly demonstrated the importance of considering skill levels of users when developing programs but also underscored the need to accomplish usability testing to ensure the final product was not hampered by the improperly matched skill level of the user.

Measuring

Virvou and Katsionis (2006) looked at children using computers both at school and at home. To measure usability at school, they used both computer logging software and self-reporting. For home use, they relied on self-reporting. They identified three characteristics of learner usability: “interface acquaintance,” “navigational effort,” and “environmental distractions” (Virvou & Katsionis, 2006, p. 163). Interface acquaintance looked at how easily the user could learn to interface with the game. Navigational effort looked at how easy it was to navigate within the simulation. Environmental distractions looked at elements in the game that could sidetrack the learner from the real point of the material. This last item could not easily be measured by self-reports since it examined cases where the user was missing the point of the lesson, which was not easy to self-identify. Blasi and Alfonso (2006) used self-reports from surveys and interviews to measure usability. Clearly usability can be measured multiple ways. One clear problem with distance learning students working on their home computers is invasive data collection in the form of observation or computer logging software is problematic at best. The distance learning format is more conducive to the use of surveys conducted either through questionnaires or interviews.

Military Studies

The military has a great deal of experience with simulators and war games. While both can encompass training and educational elements, simulators, such as flight simulators, typically focus on training. For example, a flight simulator would typically be used to practice how to handle a loss of engine power in flight and safely land the aircraft. War games generally focus more on the education. It is important to note a distinct difference between war games and computer simulations. War games are role-playing simulations carried out with large groups of people (Rubel, 2006). Typically, war games involve teams of people working as the friendly force, the enemy force, and the referee group. Computers are normally used to assist the human players; however, the emphasis is not on the computers. The emphasis is the coordination and decision-making that takes place within the friendly force group. Therefore, although a computer simulation might be used in a war game, a computer simulation by itself is not a war game. However, it is relevant to point out that while war games are different than computer simulations, the common use of war games might favorably influence military officers' perspectives on the use of role-playing computer simulations for education.

Rubel (2006) pointed out three reasons war games are used: (a) they provide insights into weakly structured problems, (b) they help gain acceptance of concepts or doctrine, and (c) they help organizations learn to interact. War is a weakly structured problem because it is inherently a dynamic situation that in its real form cannot be scripted out. Therefore, war games cannot predict outcomes of future events, but they can provide insights through the use of visualization. As Rubel pointed out,

> [War] games allow players and observers to see relationships - geographic, temporal, functional, political, and other - that would otherwise not be possible to discern. Seeing and understanding these relationships prepares the mind for decisions in a complex environment. (p.112)

Additionally, new concepts or doctrine can be explored through the use of war games to improve understanding and familiarity that can help gain acceptance of the concept or doctrine. Finally, organizations that do not typically interact but might need to in critical situations can learn how to work together more effectively by engaging in a war game together. An example would be a war game incorporating military, federal, and state organizations to deal with an emergency response to a domestic terror event. While there are clear similarities, war games address these three areas in ways computer simulations by themselves cannot.

However, war games, like computer-based role-playing games, depend on context for relevance. Schwalbe (1993) noted the war games being used at that time were based on a Cold War scenario no longer applicable in a post-Cold War era. Lack of an appropriate scenario to provide context for the war games was detrimental to the value of the game. The games needed relevant scenarios to provide the "legitimacy" players needed.

To explore the educational benefits of computer games for the military, Fong (2006) looked at using commercial games for the Singapore Air Force to assist in idea generation and experimentation. He noted this was different from the U.S. Marine Corp using *DOOM* and the U.S. Army using a game called *America's Army* because those games were designed mostly for training. Fong pointed out the advances in technology are enabling games and military simulations to merge. If they are created properly, simulation can be engaging enough for game savvy soldiers that it would then significantly decrease training required. Variability of player proficiency was noted. Fong recommended specifically using role-playing games for education because those games are not time critical, which helps overcome the disadvantage inexperienced users have with time sensitive games. Use of games for idea generation was deemed very beneficial and successful.

Although the use of robust computer role-playing simulations for Air Force distance learning PME is a relatively new and unstudied phenomenon, there have been studies of Air Force PME itself and those studies merit examination to better understand the PME program. Before examining those studies, it is worth pointing out problems with metrics for evaluating online PME programs. In 2004, the Government Accounting Office (GAO) found a lack of relevant metrics for evaluating PME programs that utilize newly distributed learning elements to augment or replace more traditional correspondence programs. Although standard identified objectives for all PME programs are listed in the Officer Professional Military Education Policy (OPMEP), no metrics are specified in the OPMEP document for evaluating this new approach. The GAO also did not provide any recommendations for appropriate metrics in their report.

Two other relevant studies examined Air Force PME. The first was accomplished by Kraska and Bentley (2004) who looked at the entry level officer PME program called the Air and Space Basic Course (ABSC). They examined the graduates' perceptions of relevance and effectiveness of curriculum and compared them to the supervisors' perceptions. Their findings noted the supervisors actually perceived a higher level of effectiveness from the PME program than did the graduates. Kraska and Bentley found the difference to be significant but they were unable to determine the cause of the disparity. The possible causes they discussed included the idea that supervisors might be in a better position to compare ABSC graduates and non-graduates, and supervisors were expecting to see improvement, which might have altered their perception (Kraska & Bentley, 2004). This study did not answer the question of who was in the best position to judge PME effectiveness.

A second study was conducted by MacCuish in 2001 that examined the intermediate level Air Force PME program--Air Command and Staff College (ACSC) distance learning program. This was conducted as a status study intended to report on the current conditions of the ACSC distance learning program. The ACSC program at the time was similar to the Air War College program today with the exception of the cultural simulation program. The study noted problems in curriculum development, the need for better multimedia material, and the limited student assessments. The limited assessments problem was in line with the GAO findings of 2004 that found the distance learning program was only able to evaluate at the comprehension level due to the nature of the testing process. It is important to note the in-residence program taught and tested up to the synthesis level.

Summary

The multitude of advantages simulations and games offer for education, coupled with their strong pedagogical underpinnings, justify consideration for use in education. The significant advancement of computers and the associated ability to create more robust simulations have produced an opportunity for education to greatly expand the use of simulations and games to better engage students in the process of learning. However, as the review of literature demonstrated, educational simulations and games must be intentionally developed with a firm understanding of the pedagogical implications as well as a solid understanding of how the student's background impact the simulation experience in order for the programs to be effective learning tools. Although there have been a multitude of studies, the body of knowledge is still incomplete. In particular, the impact gender as well as experience and attitudes toward video games might have on effectiveness, motivation, and usability of simulations needs to be more thoroughly examined. The relative dearth of Air Force PME studies and the recent addition of a robust cultural simulation program to the PME curriculum also justified further research in this area.

CHAPTER III

METHODOLOGY

Introduction

This chapter describes the methodology used to complete the study. Air War College (AWC) simulation programs, population, Institutional Review Board (IRB) approval, and study sample are detailed. Next, the variables are discussed, the instrument is described, and additional sources of data are explained. Finally, the design and statistical plan used for conducting the survey research on the AWC cultural simulation program is covered.

Overview of Air War College Cultural Simulation Programs

Air War College is designed to provide Lieutenant Colonels with professional military education to prepare them for future challenges they will face in their careers. Completion of AWC is not specifically required by regulation. However, it is very rare to be promoted to Colonel without completing the AWC program, effectively making it mandatory for Air Force officers to be competitive for promotion. The in-residence AWC program awards a master's degree but the distance learning program only provides a certificate of completion. The distance learning program covers the same material as the in-residence program but the curriculum is not as in-depth. There is very little interaction with the instructor and no interaction with other students so awarding the master's degree is not justifiable. Based on Secretary of the Air Force Michael Wynne's (2007) guidance to increase cultural skills, a cultural understanding course was added to AWC distance learning curriculum in January 2008. This course utilized a role-playing cultural simulation program designed to help the students examine the practical application of cultural understanding. A second culture simulation program became available and was added to the curriculum during the data collection phase of this research. The second simulation was Web-based and provided an alternative for students who experienced technical problems installing the original simulation on their computer. Students were required to successfully complete one of the two simulations in order to pass the course. Completing the cultural simulation was typically the last activity students accomplished prior to graduating from Air War College. Data were collected from students who had completed one of these simulation programs.

Original Simulation

The original simulation program was set in the context of a humanitarian relief operation in Africa. There were seven scenarios within the simulation. Each scenario involved multiple animated scenes requiring the student to make decisions on how to interact with the various characters to accomplish the mission. Figure 1 shows an example of the animated presentation experienced while running the simulation. The simulation used multiple branches so decisions made by participants during each scenario impacted the subsequent choices available as well as how the characters interacted later in the simulation. The participants were all given the same initial scenario and the same clearly defined overall mission, but there were multiple paths to complete the simulation. The decisions made by a student impacted the experience within each scenario and also determined the number of scenarios required to successfully complete the program. The student needed to deal with various problems during each of the scenarios while ensuring continued progress toward the overall objective. When students actions resulted in the overall mission becoming unobtainable or the student successfully completed a scenario in the simulation, a debriefing was provided. The debriefing was text-based feedback listing how the user did in each of the scenarios. This feedback was generic in nature. If the student failed the mission, the feedback identified the overall problem but it did not point out which specific decisions led to the failure. The intention was to force students to think through their actions and determine what they should do differently. Additionally, when a student failed the mission, a “professor” character was available on the feedback screen to provide a short audio "lecture" about a variety of general cultural topics such as universalistic versus particularistic. The program had to be downloaded and installed on the student's home or government computer. The amount of time needed to complete the program depended on the student's actions but it generally took a few hours to complete. When the simulation was completed, the software provided the student with a completion code to submit to Air War College. This provided proof of completion but also contained the student's identification number and the choices the student made to complete the program.

Figure 1. Screen shot of original simulation animated interface.

Visual Expeditionary Skills Training Simulation

As previously noted, partway through data collection, a new simulation was offered as an alternative to completing the original simulation. Students had the choice to use either simulation to complete their Air War College requirement. The newer simulation program, entitled visual expeditionary skills training (VEST), was created by the Air Force Culture and Language Center to prepare personnel for deployment to either Afghanistan or Iraq. The simulation was completed and became available in April 2011. This simulation program utilized live actors to present an immersive experience in dealing with cultural situations the user might actually encounter. The program was Web-based and not installed on the user's computer. Some students had been unable to install the OS simulation on their computer so the Web-based VEST simulation provided a viable alternative to accomplish the culture requirement. As with the original simulation, the student faced a series of choices. As the background in Figure 2 depicts, the simulation incorporated video footage with actors to provide a more life-like presentation than the original simulation. The center of Figure 2 presents an example of choices the user faced during the simulation. Unlike the original simulation, VEST was a "railroad" type simulation that only had one path to completion and forced a student to "get back on track" to progress. Each time a student made a choice, feedback was given and if the choice was less than desirable, the simulation was reset to the decision point and the student was able to make a different choice. Once the correct choice was made, the simulation progressed to the next decision point. Every choice resulted in feedback via video footage showing what would have happened if their chosen option was implemented. This was intended to help the user better understand the effect of that choice. Students might have viewed additional video footage depending on their choices but overall, the students completing the VEST simulation were all faced with the same scenarios and options. Upon completing the program, the student took a test that was not embedded within the program. The test was located on the Distance Learning Blackboard website. The program was added as an option to students in June 2011

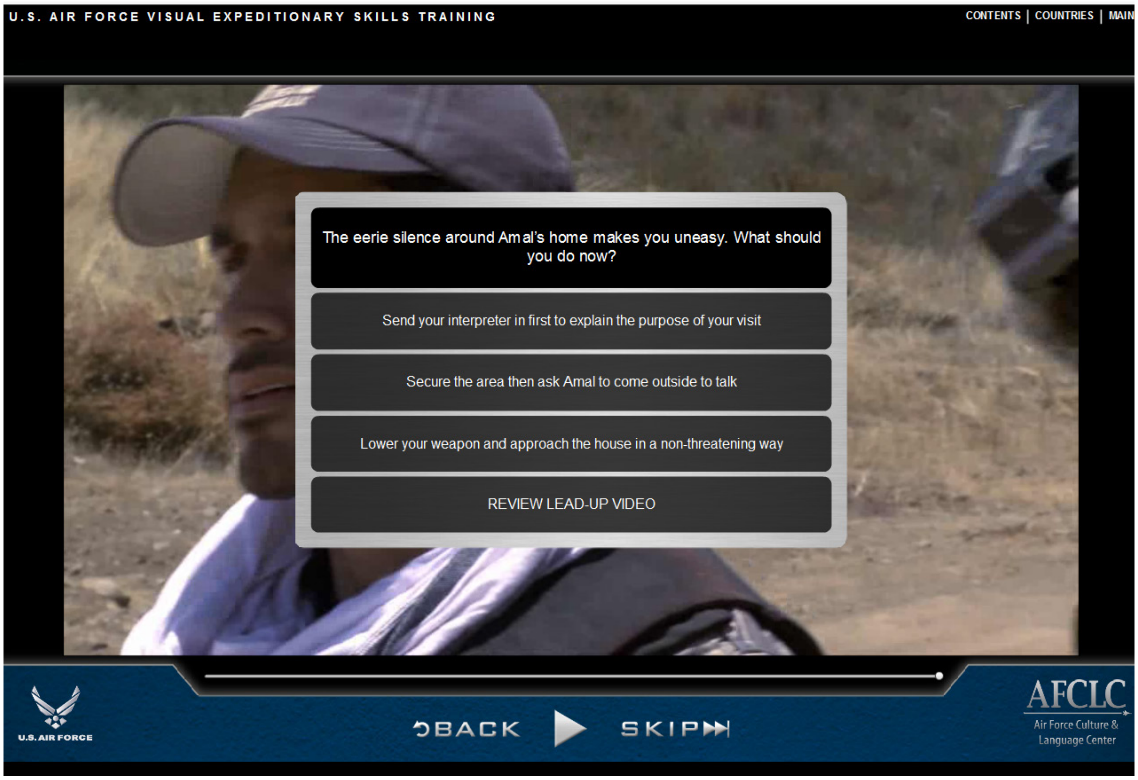


Figure 2. Screen shot of visual expeditionary skills training interface.

Population

The target population for this study was military officers, generally at the Lieutenant Colonel rank, who normally had 15 to 20 years of military experience. They were generally in their mid-30s through mid-40s. They all had a bachelor's degree and nearly all had a master's degree as well. Roughly 13% of this population was female (2011 USAF Almanac, 2011).

The accessible population consisted of Air War College Distance Learning (AWC/DL) students who had accomplished the cultural simulation within the AWC/DL curriculum. The AWC/DL program was voluntary but the promotion rate for officers who failed to complete the course was nearly 0%. Therefore, although students volunteered for the program, they had a variety of reasons and motivation levels for completing the AWC/DL program.

Sample

Initially, the plan was to send the survey to everyone who had graduated from the Air War College Distance Learning (AWC/DL) program in the previous year. However, Air Force rules at the time necessitated sending out surveys only to those students who had completed their graduation requirements in the previous week. This change increased the length of time needed for data collection but also potentially benefited the study by ensuring nearly all the survey recipients had recently completed the simulation. The cultural simulation course was normally the last requirement accomplished prior to graduation. In those cases where it was not the last requirement completed, there were normally only a few weeks between the completion of the simulation and program completion. The proximity of completing the simulation helped ensure the simulation was still recent enough in the student's memory for him or her to reasonably assess the experience. Based on that change, from September 29, 2010 to March 28, 2012, surveys were sent each week to all AWC students who completed AWC/DL program requirements during the previous week. A target of at least 250 returned surveys was initially set to ensure the number of female respondents would be sufficiently large enough for statistical analysis. When the new simulation was added, the collection was extended until the completed surveys for the new simulation had an adequate sample size to be analyzed and compared. The final target was 250 completed surveys for the VEST simulation. Based on approximately 13% of the accessible population being female, this was intended to provide roughly 30 female participant surveys for the new simulation.

Variables

The independent variables for this research study were prior gaming experience and gender. The dependent variables were perceived effectiveness, usability, and Keller's (2010) motivation categories of attention, relevance, confidence, and satisfaction. Perceived effectiveness measured students' perception of their ability to apply the knowledge they gained from the cultural simulation program. Usability measured student perceptions of their experience interacting with the simulation in terms of how easy or difficult it was to use the simulation. Keller and Suzuki (2004) explained the four categories of motivation, noting that attention referred to the need to acquire and then maintain the student's attention. Maintaining attention included using variety so the learner does not become bored with a predictable approach. Relevance referred to the student's impression of the material being linked to their job, personal aspirations, and experiences. Confidence referred to the student expecting to successfully complete the material and believing their success was dependent on their actions and not simply on luck. Satisfaction referred to creating a positive reaction to the experience. Satisfaction included the students viewing the value of knowledge gained was worth efforts expended to accomplish the simulation (Keller & Suzuki, 2004).

Instrument

The instrument for this research was a 53-question survey conducted via a Web-based survey system. A copy of the survey is provided in Appendix A. The survey included 10 multiple choice questions covering confirmation of simulation completion, time needed to complete the program, gender, age, military service, rating (flying or support career field), prior wargame experience, and background on the subject's attitude toward and recent experience with playing video games. The next three questions addressed effectiveness and usability. These three questions used an 8-point Likert scale answer system ranging from *Strongly disagree* to *Strongly agree*. Perceived effectiveness looked at students' perceived ability to take what they had learned in the simulation and apply it in the workplace. Usability looked at two of Virvou and Katsionis's (2006) characteristics of usability, interface acquaintance, and navigational effort. The next 36 questions were adapted from Keller's (2010) instructional materials motivation survey (IMMS) and covered Keller's four areas of motivation: attention, relevance, confidence, and satisfaction. These questions were rated on a 5-point Likert scale ranging from *Not true* to *Very true*. The original wording of Keller's survey was changed based on his instructions for customizing the survey without changing the characteristics of the survey results. The last two questions were open-ended questions dealing with what the students liked most and least about the simulation. Those two questions provided a way for students to further articulate their impressions of the cultural simulation program and the use of simulations in general. At the end of the survey, students were given the option to volunteer for follow-up interviews if needed.

Although the IMMS portion of this instrument was previously used, the rest of the survey was not so steps were taken to examine validity and reliability for the instrument. The non-IMMS portion of the survey was assessed for face validity by having the Dean of Air War College Distance Learning and two of his instructors review it relative to the operational definitions of effectiveness and usability to ensure the instrument was measuring the two variables and nothing else. A pre-sample using AWC/DL instructors was then used to help ensure the presentation of the survey was not confusing or misleading. Additionally, although the survey was sent out weekly for 18 months, the returned surveys from the first six weeks were examined to ensure the notification system worked properly and the students were able to utilize the website to complete the survey. This provided an opportunity to adjust the survey if needed. This early inspection of the data included using Cronbach's alpha to examine the internal reliability of the survey. No major changes to the survey were required and student responses from the first six weeks were combined with subsequent surveys. The only change made to the survey after the initial launch was to the first question. This was done to accommodate the new simulation (VEST) that was added to the AWC/DL program approximately nine months after the data collection had begun. Instead of asking if they completed the simulation, question one was changed to ask if they completed the original simulation, VEST, or did not complete either simulation. This change was considered minor and unlikely to cause any change in response rates or rating levels.

The IMMS portion of the survey had a great deal of prior research to establish its reliability and validity. Reliability estimates for the IMMS were based on using Cronbach's alpha and were broken into the four areas of motivation it assessed. Previous research provided a reliability estimate of 0.89 for attention, 0.81 for relevance, 0.90 for confidence, 0.92 for satisfaction, and 0.96 for the entire survey.

Additional Sources of Data

In addition to the survey, two additional sources of data were used: completion codes for the simulation and demographic information. Both sources were already compiled by AWC.

Completion codes for the original culture simulation were transmitted to AWC by each student as proof they successfully completed the simulation. These codes contained information on students' choices during their final successful attempt to complete the simulation. These codes had been previously collected by AWC but prior to this study, they had not yet been examined by anyone. The codes only contained information on the choices the student made to successfully complete the simulation and, therefore, did not provide information on unsuccessful attempts. This was a limitation of the data but it still provided a potential source of insight into general use of the simulation, student decision-making, and patterns of choices linked to the other factors.

The AWC provided limited demographic information. First, they provided the gender for each of the completion codes. Second, they provided email addresses for students who completed the simulation. The email list simply identified the graduating student's email address without providing any other identifying or sensitive information.

Design

A list of student email addresses was supplied weekly by AWC for students who had completed all of their AWC/DL requirements during the previous week. An email was sent to each of those students inviting them to participate in the voluntary Web-based survey. The results from the first six weeks were examined to identify if changes were needed. The first area to be checked was whether the email generated by the survey system was working properly. This was considered an essential step because many of the students were deployed and it was necessary to determine if there were issues with access to the survey website from deployed locations. Additionally, the first six weeks of surveys were analyzed as an initial check of reliability to determine if adjustments needed to be made to the survey wording. If changes were needed, adjustments would have been made and the new survey would have been sent to subsequent groups of students. Changes minor in nature would have allowed the initial data to be combined with subsequent surveys. If large changes were needed, the groups would have been treated separately. Emails for both groups covered everyone who completed the program within the time window of September 29, 2010 to March 28, 2012. This time window resulted in sending emails to a total of 2,104 individuals. The participants responded by logging onto the website and completing the survey. The survey window for each group was approximately 45 days with a reminder email sent out after two weeks. The AWC also supplied data on student completion codes and some demographic information. The resulting data were then analyzed. If follow-up interviews were deemed necessary to fully explain the results, the interviews would have been conducted with those students who volunteered for follow-up interviews in their survey responses.

Statistical Plan

The two independent variables were video game experience and gender. These were categorical variables. Video game experience was handled by combining the responses from the two survey questions dealing with video game experience and using a median split to separate respondents into a group with more experience and a group with less experience. The dependent variables were based on Likert scale responses. Those responses were treated as though they were on an interval scale. As noted by Creswell (2008), this is a debatable approach but has become common practice and the risk to the analysis involved with treating it as interval data is low. This assumption allowed the use of parametric statistical tests. The SPSS software was used for all statistical analysis for this research.

Five research questions were examined using the methods covered in the following paragraphs. Questions 2, 3, and 4 were grouped together for discussion purposes because they utilized the same statistical method.

Research Question 1

Is the cultural simulation program perceived as effective?

As an instruction course designed by professional educators, the program was hypothesized to be effective. The question was examined using descriptive statistics utilizing the responses from survey question 11. This survey question asked if the subject felt the knowledge gained from the AWC cultural simulation program helped prepare them for dealing with cultural issues in future assignments. This directly connected the student's perception to the purpose of the course and whether he/she perceived the simulation to be effective in providing usable knowledge. The result was a self-reported assessment of whether the simulator was perceived as effective. The scale for this question ran from a 1--*Strongly Disagree* to 8--*Strongly Agree*. A mean above 4.5 would indicate the simulation was perceived as effective and a mean below 4.5 would indicate the simulation was not perceived as effective.

Research Questions 2, 3, and 4

The second, third, and fourth research questions all looked at the same dependent variables but through different independent variables or interactions. The three questions are briefly described below and then the statistical plan to explore those questions is covered.

Research question 2 asked:

> Does prior gaming experience impact perceived effectiveness, usability, attention, relevance, confidence and satisfaction when using role-playing simulations for Professional Military Education (PME)?

Prior gaming experience was hypothesized to improve perceived effectiveness, usability, and the four aspects of motivation.

Research question 3 asked:

> Does gender impact perceived effectiveness, usability, attention, relevance, confidence and satisfaction when using role playing simulations for PME?

Literature on the impact of gender was mixed with some studies finding decreases in effectiveness, usability, and motivation in using simulations and others finding no difference. However, unlike those studies, female military officers were not expected to be truly representative of the non-military female population as a whole in terms of attitudes toward technology. These officers self-selected to be in high tech career fields and, therefore, it was hypothesized there would be no significant impact on effectiveness, usability, and the four aspects of motivation due to gender.

Research question 4 asked:

> Is there an interaction between gender and gaming experience on perceived effectiveness, usability, attention, relevance, confidence and satisfaction when using role-playing simulations for PME?

Since no significant difference was expected based on gender, the hypothesis was there would be no significant interaction between gender and gaming experience on perceived effectiveness, usability, and motivation. However, even if there was no effect noted by gaming experience or gender, there could be a difference in the interaction of the two variables due to a disordinal interaction, which could mask the effects of video game experience and gender.

Research questions 2, 3, and 4 were analyzed using the survey data. A factor analysis was utilized to confirm if a four-factor structure existed for the survey's motivation questions. This confirmatory process was intended to help support the value of the subsequent analysis on each of the four aspects of motivation. Following the factor analysis, the survey data were analyzed using a two-way multivariate analysis of variance (MANOVA). This compared the two independent variables--video game experience and gender--with the dependent variables of effectiveness, usability, attention, relevance, confidence, and satisfaction. An alpha of .05 was used. Based on the results, if a significant MANOVA statistic existed, then a stepwise discriminate analysis was used to determine which dependent variables were most responsible for the variation (Heiny & Mundfrom, 2010). Additionally, numbers, means, and standard deviations for each question were computed.

Research Question 5

Research question 5 asked:

What aspects of the simulation do participants find valuable or problematic? Question five was answered using data from the two open-ended questions. These questions were reviewed for trends or other issues not represented by the other survey questions. This involved a qualitative assessment and the resulting themes had the option of being explored as needed with follow-up interviews.

Additional Analysis

Completion codes from the original simulation were analyzed using a chi-square statistic to determine if males and females made different choices during the simulation.

Additionally, surveys from the new simulation were analyzed using the same procedures described above for the original simulation.

Institutional Review Board

This study was approved by the Institutional Review Board at the University of Northern Colorado in May 2010 (see Appendix B).

CHAPTER IV

RESULTS

Introduction

This chapter describes and discusses the results of the data analysis. First, a description of the survey implementation and the initial six-week data analysis is covered along with survey response rates and demographics. Next is a discussion on how incomplete data in the surveys were treated, how video categories were created, and an analysis of internal reliability. Then the research questions are covered as grouped in the previous chapter including a discussion of the results for each question. For each question, the original simulation dataset analysis is covered followed by the VEST dataset analysis. Finally, two additional data analyses are covered--first the completion code analysis and then a post hoc analysis using the participant's age as the independent variable.

Implementation and Six-Week Analysis

The survey was sent out weekly starting on October 5, 2010 to all students who had completed the program requirements the previous week. This continued weekly until March 28, 2012. The first six weeks of student responses were analyzed to accomplish an initial check of the survey delivery system and to check the usability and reliability of the instrument. Additionally, student responses were reviewed to determine if any problems were noted in the open-ended comments. Seventy-three survey invitations were sent out during the first six weeks. Two students marked they did not want to take the survey. One student only partially completed the survey, 51 students fully completed the survey, and the remaining 19 never responded. Overall, the completion rate for this group of students was 70.7%. Data for the partially completed survey were discarded because the student left 19 questions blank. The point at which the incomplete survey was discontinued was in the middle of the questions from Keller's (2010) Instructional Materials Motivation Survey (IMMS) so there was no reason to believe the stopping point was the result of the wording of any particular question. Additionally, no written comments were included in the open-ended question of that partially completed survey.

Four of the 51 surveys were from female participants. This accounted for 7.8% of the responses. The target population of lieutenant colonels contained approximately 13% females so this was less than expected. However, this was based on a very low number of surveys. There was no way to determine the demographics of the survey invitees so it was not considered indicative of a gender issue with the invitations or the survey.

A Cronbach's alpha was computed for each applicable variable for this initial group of completed surveys. As shown in Table 1, the "usability" questions had a Cronbach's alpha of 0.645, which was the lowest. Since the *N* of 51 was low, the Cronbach's alpha was expected to increase with a larger *N*. Although Keller's (2010) IMMS questions had been used extensively and had a well-established reliability, the Cronbach's alphas were also generated for those questions to help ensure the slight modifications made to tailor the questions for use on this survey instrument did not result in any unintended confusion or alteration of the intent of the question. The Cronbach's alpha for each of the variables was deemed acceptable so no changes were made to those questions.

Table 1

Cronbach's Alpha for First Six Weeks of Surveys

Category	Number of Questions	Cronbach's Alpha
Video Game Experience	2	0.867
Usability	2	0.645
Attention	12	0.923
Relevance	9	0.862
Confidence	9	0.723
Satisfaction	6	0.903

The open-ended questions were also examined to determine if the survey instrument was identified in the comments as being problematic or if students were unsure about what was being asked. Forty-three of the 51 participants completed the open-ended question regarding what they liked most about the simulation and 45 completed the open-ended question regarding what they liked least. The answers provided specific thoughts about the simulation, which are covered later in this chapter along with comments from all subsequent survey responses. No answers from the open-ended question mentioned any difficulty with the survey instrument itself.

There were email messages and phone calls from many respondents to the researcher during this time period regarding the survey. The Air Force annually trains its service members on cyber-threats and computer security. These emails and phone calls all dealt with verifying the survey was legitimate and not an attempt to breach security firewalls. In addition to the civilian email address already listed in the email invitation, the researcher's military email address was added to the invitation email to make it easier for invitees to make contact with the researcher and ensure these concerns of legitimacy could be readily addressed. However, it is important to acknowledge there might have been students who simply deleted the email rather than make an effort to try to identify the authenticity of the survey.

This initial analysis was conducted to identify if changes were required to the survey. A change was made to the initial invitation email providing the military email address in addition to the civilian email address of the researcher. However, the survey itself was not changed at this time. Approximately nine months after the survey collection began, a change was made to the first question to deal with the inclusion of a second simulation to the AWC/DL program.

Survey Response Rates and Demographics

The last set of survey invitations was sent on March 28, 2012. Data collection was concluded on May 15, 2012. From September 29, 2010 to March 28, 2012, a total of 2,104 survey invitations were sent to students who had completed the culture simulation. Approximately 56.7% of the invitees responded for a total of 1,192 returned surveys. There were 167 surveys that had at least one multiple-choice answer left blank. Twenty-nine of the incomplete surveys were removed and 138 were kept. An explanation of how partially completed surveys were dealt with is contained in the next section. The demographics of the returned survey results provided in Table 2 show breakouts by gender as well as by the type of simulation completed. Approximately 11.8% of the surveys were completed by females. This compared to roughly 13% of the lieutenant colonel population being female. Of the surveys used for analysis, 898 surveys were based on the original simulation and 265 were based on the newer VEST simulation.

Table 2

Survey Return Results

Results	*N*
Survey Invitations Sent	2,104
Survey Responses	1,192
Incomplete Surveys	167
Incomplete surveys removed	29
Incomplete surveys retained	138
Total number of surveys used for analysis	1,163
Original Simulation Total for analysis	898
Original Simulation - Males	792
Original Simulation - Females	106
VEST (New Simulation) Total for analysis	265
VEST - Males	230
VEST - Females	35

Treatment of Incomplete Data

The 167 incomplete surveys were evaluated to determine if the results could be kept as part of the database for statistical analysis or if they needed to be removed. In all cases, the open-ended responses were retained. If more than four of the 49 multiple-choice questions were blank, then that survey was removed. The cutoff of four questions was selected because it represented approximately 10% of the total questions and the data indicated it was a logical tapering point. This resulted in 29 surveys being removed from the statistical data base. A breakout of the number of questions left blank on the incomplete surveys is listed in Table 3.

Table 3

Incomplete Survey Results

Number of Blanks	All Surveys	Original Simulation	VEST
1 blank question	116	88	28
2 blank questions	14	7	7
3 blank questions	6	5	1
4 blank questions	2	2	0
6 blank questions	1	1	0
7 blank questions	3	3	0
8 blank questions	1	0	1
13 or more blank	24	20	4
Total incomplete surveys	167	126	41

The 138 incomplete surveys that had four or less blank questions were retained in the database with adjustments using the following rules. Questions that were not part of the Keller (2010) IMMS were filled in with survey averages for that question from the completed surveys. If the question was from Keller's IMMS, then the average of that individual survey's responses was used. For example, there were 12 questions on attention in the IMMS portion of the survey. If one of the 12 attention questions was blank, the average of the other 11 was used to fill in the blank. A total of 170 blanks were filled in using these rules. Thirty-eight blanks were on non-Keller questions and 132 were on questions from Keller's IMMS. These rules minimized the impact on the dataset while retaining 6,592 inputs from the filled in questions on these 138 surveys. Although the blanks were scattered throughout the survey, the very last multiple-choice question was left blank 16 times. The survey page listed the question at the top of the page just before the open-ended questions, which might have distracted the respondents and caused them to focus on the open-ended questions while missing the final multiple-choice question.

Categorizing Video Game Experience

Two Likert-scale questions dealt with the independent variable of video game experience (VGE). For statistical analysis, the 1,163 survey participants had to be categorized as experienced video game players or inexperienced video game players. To accomplish this, the numerical results of the two Likert-scale questions were added together and then a median split was used to divide the participants into the two groups. This was accomplished separately for male and females.

As shown in Table 4, the median split for males was between four and five. There were 573 males who rated themselves with a net VGE of four or less and were categorized as inexperienced video gamers. There were 449 males who rated themselves with a five or above and were categorized as experienced. The median split for females was between three and four as shown in Table 5. Females rating themselves with a net VGE of three or less were categorized as inexperienced and ratings of four or above were categorized as experienced. This resulted in 76 inexperienced female video game players and 65 experienced female video game players.

Table 4

Video Game Experience Split for Males

Net VGE	# of Males	Split
2	88	
3	161	573
4	324	
5	178	
6	174	449
7	52	
8	45	
Total	1022	

Table 5

Video Game Experience Split for Females

Net VGE	# of Females	Split
2	40	76
3	36	
4	48	65
5	8	
6	6	
7	0	
8	3	
Total	141	

Internal Reliability

A Cronbach's alpha was computed for variables that had multiple questions on the survey. The results are listed in Table 6. Usability had a Cronbach's alpha value of .707 for this dataset, which was higher than the six-week analysis. However, similar to the six-week analysis, usability was still the lowest Cronbach's alpha value. Each of Keller's (2010) IMMS categories had a previously established Cronbach's alpha. Except for the confidence variable, there were only slight differences between the previously established values calculated from this survey dataset.

Table 6

Cronbach's Alpha Results For All Surveys

Category	Number of Questions	Cronbach's Alpha	Prior Established Cronbach's Alpha.
Video Game Experience	2	0.827	N/A
Usability	2	0.707	N/A
Attention	12	0.926	.89
Relevance	9	0.866	.81
Confidence	9	0.790	.90
Satisfaction	6	0.896	.92
All Motivation (ARCS)	36	0.959	.96

Descriptive Statistics

Descriptive statistics for the original simulation are presented in Table 7. The means for the independent variable of gender did not appear to have a pattern. However, the means for VGE did appear to have a distinct pattern with means for inexperience video gamers being lower for all six variables. A more thorough examination of these means is provided later in this chapter utilizing a MANOVA analysis.

Table 7

Descriptive Statistics for the Original Simulation

Background	Sample size (N)	Perceived Effectiveness $\bar{x}$ (S.D.)	Usability $\bar{x}$ (S.D.)	Attention $\bar{x}$ (S.D.)	Relevance $\bar{x}$ (S.D.)	Confidence $\bar{x}$ (S.D.)	Satisfaction $\bar{x}$ (S.D.)
Gender							
Male	792	4.040 (1.822)	8.525 (3.393)	38.845 (10.055)	27.155 (7.460)	33.448 (5.653)	15.444 (5.883)
Female	106	4.103 (1.831)	7.632 (3.563)	40.500 (9.120)	27.123 (6.700)	33.500 (5.269)	15.500 (5.571)
VGE							
Inexperienced	501	3.896 (1.798)	7.968 (3.466)	38.575 (9.891)	26.920 (7.452)	32.780 (5.788)	15.080 (5.700)
Experienced	397	4.239 (1.836)	8.990 (3.286)	39.627 (10.025)	27.443 (7.266)	34.305 (5.253)	15.451 (5.844)
Time to Completion							
<2 hours	31	3.387 (1.856)	9.839 (2.782)	37.323 (11.185)	24.581 (8.725)	34.323 (5.546)	13.387 (5.998)
2 to <4 hours	248	4.016 (1.812)	8.685 (3.479)	38.621 (10.301)	26.927 (7.398)	34.524 (5.132)	15.206 (5.967)
4 to <6 hours	303	4.129 (1.792)	8.620 (3.1973)	39.429 (9.436)	27.568 (7.040)	33.743 (5.222)	15.663 (5.763)
6 or more hours	316	4.060 (1.822)	7.880 (3.573)	39.165 (10.067)	27.180 (7.495)	32.253 (6.105)	15.642 (5.788)
Age							
30-35	13	4.231 (1.739)	8.692 (3.449)	38.769 (7.991)	26.077 (6.922)	33.692 (4.626)	14.538 (5.109)
36-40	380	3.713 (1.819)	8.500 (3.396)	37.797 (9.928)	25.982 (7.480)	33.329 (5.488)	14.211 (5.693)
41-45	314	4.089 (1.744)	8.475 (3.458)	38.987 (9.630)	27.567 (7.126)	33.697 (5.571)	15.519 (5.586)
46-50	136	4.507 (1.858)	8.140 (3.349)	40.801 (10.536)	28.596 (7.250)	33.140 (6.251)	17.588 (6.063)
>50	55	4.945 (1.682)	8.182 (3.667)	43.636 (9.278)	29.545 (7.162)	33.454 (5.257)	18.564 (5.367)

Table 7 Continued

Background	Sample size (N)	Perceived Effectiveness $\bar{x}$ (S.D.)	Usability $\bar{x}$ (S.D.)	Attention $\bar{x}$ (S.D.)	Relevance $\bar{x}$ (S.D.)	Confidence $\bar{x}$ (S.D.)	Satisfaction $\bar{x}$ (S.D.)
Branch of Service							
Air Force	832	4.004 (1.824)	8.379 (3.422)	38.831 (9.921)	27.020 (7.294)	33.371 (5.628)	15.246 (5.793)
Army	11	4.727 (1.7939)	10.091 (2.845)	42.455 (11.148)	31.364 (9.341)	34.909 (4.549)	19.000 (5.882)
Navy	2	6.500 (0.707)	10.500 (2.121)	51.000 (9.900)	39.000 (7.071)	36.500 (6.364)	23.000 (8.485)
Marines	15	4.867 (1.552)	8.467 (4.033)	44.400 (9.440)	32.400 (6.197)	35.533 (5.181)	18.400 (6.390)
International	1	6.000 (0.000)	13.000 (0.000)	49.000 (0.000)	27.000 (0.000)	37.000 (0.000)	18.000 (0.000)
Civilian	37	4.320 (1.749)	8.595 (3.362)	39.649 (9.959)	26.081 (7.661)	33.784 (5.588)	15.451 (5.844)
Type of Service							
Active Duty	628	3.885 (1.820)	8.455 (3.426)	38.279 (9.867)	26.787 (7.460)	33.416 (5.635)	14.793 (5.718)
Reserve	137	4.620 (1.783)	8.657 (3.381)	41.891 (9.797)	28.759 (7.120)	33.905 (5.354)	17.460 (5.980)
Guard	82	4.207 (1.705)	7.817 (3.297)	39.695 (10.213)	27.598 (6.818)	32.927 (5.788)	16.146 (5.802)
Civilian	51	4.255 (1.820)	8.314 (3.696)	39.706 (9.878)	26.608 (7.357)	33.569 (5.697)	17.039 (5.568)
Occupation Rated	310	3.561 (1.686)	8.094 (3.218)	36.629 (9.169)	25.371 (7.003)	32.900 (5.429)	13.632 (5.267)
Non-rated	588	4.304 (1.840)	8.592 (3.518)	40.311 (10.130)	28.090 (7.392)	33.747 (5.681)	16.410 (5.844)
Wargame Experience							
Yes	706	3.994 (1.814)	8.455 (3.423)	38.955 (10.139)	27.180 (7.565)	33.531 (5.686)	15.246 (5.879)
No	192	4.245 (1.841)	8.292 (3.432)	39.354 (9.283)	27.047 (6.626)	33.172 (5.308)	16.203 (5.844)

Among the four Time to Complete categories, those participants who spent less than two hours completing the simulation had the lowest means for perceived effectiveness, attention, relevance, and satisfaction. In fact, the means of those who spent less than two hours had the lowest means out of all the background categories for perceived effectiveness, relevance, and satisfaction. That group also accounted for the highest mean of any category for the usability variable. The "six hour or more" category accounted for the lowest mean for confidence of any background category.

For the Age of Participant categories, there appeared to be a pattern with perceived effectiveness, attention, relevance, and satisfaction ratings all being higher for older participants. In fact, the "greater than 50 years old" category accounted for the highest means of any category for those variables. However, the highest means for age in usability and confidence were in the 30-35 age group. The clear patterns within the age variable merited further examination. Although not part of the original research plan, a post hoc analysis utilizing a one-way MANOVA with age as the independent variable was accomplished and is presented at the end of this chapter.

Among the other background categories, Branch of Service showed the participants were primarily Air Force with very small numbers of participants from the other branches of service. Air Force respondents had the smallest means for everything except relevance. Type of Service appeared to have a pattern with Reserve Officers giving the highest mean ratings for every variable. Occupation seemed to show a pattern with non-rated participants having the highest mean for every variable. Finally, there was no noticeable pattern in Wargame Experience.

The original research plan was written prior to the VEST simulation being added to AWC/DL curriculum. However, the addition of VEST during the survey sampling of graduates afforded an opportunity to garner more insight into educational simulations for professional military education. Visual expeditionary skills training was web-based rather than a program that needed to be loaded on the participant's computer so the technical issues were different than the original simulation. Additionally, the simulation design for VEST is a "railroad" design--if a poor choice is made, then the participant is put "back on track." This is accomplished by providing immediate feedback followed by an opportunity to make a different choice for that decision point. Throughout remaining sections in this chapter, the original simulation dataset is analyzed followed by an analysis of the VEST dataset.

The descriptive statistics for the VEST simulation survey data are presented in Table 8. The data appear to show a consistent pattern of lower means for males and lower means for experienced gamers except for perceived effectiveness. The means for those completing the simulation in less than two hours were the lowest and those taking four or more hours had the highest means. Like the original simulation dataset, age appeared to have distinct trends for perceived effectiveness, attention, relevance, and satisfaction. This was explored further with a post hoc MANOVA which is presented at the end of this chapter. Type of Service did not show a distinct pattern but rated personnel did show a consistent pattern of lower mean ratings compared to non-rated personnel, which matched the results of the original simulation. Those with wargame experience had lower mean ratings for all categories except confidence. Additionally, the standard deviations for VEST were generally smaller for the gender and video game experience categories with the exception of the satisfaction variable that had a slightly higher standard deviation for all four groups with the VEST data.

Table 8

Descriptive Statistics for Visual Expeditionary Skills Training

Background	Sample size (N)	Perceived Effectivenes s $\bar{x}$ (S.D.)	Usability $\bar{x}$ (S.D.)	Attention $\bar{x}$ (S.D.)	Relevance $\bar{x}$ (S.D.)	Confidence $\bar{x}$ (S.D.)	Satisfaction $\bar{x}$ (S.D.)
Gender							
Male	230	5.583	11.087	48.113	33.470	39.465	19.652
		(1.648)	(2.890)	(8.898)	(6.025)	(3.717)	(5.980)
Female	35	5.800	11.229	50.257	35.600	39.914	21.314
		(1.530)	(3.172)	(7.868)	(7.257)	(3.501)	(5.754)
VGE							
Inexperienced	148	5.595	11.122	48.838	34.236	39.757	20.095
		(1.616)	(3.097)	(8.352)	(5.546)	(3.632)	(5.704)
Experienced	117	5.632	11.085	47.838	33.137	39.231	19.590
		(1.659)	(2.699)	(9.311)	(6.974)	(3.747)	(6.297)
Time to Completion							
<2 hours	44	4.523	10.341	42.318	29.682	38.545	16.250
		(2.029)	(3.894)	(11.379)	(7.097)	(4.401)	(5.812)
2 to <4 hours	124	5.597	11.306	49.081	34.218	39.516	20.024
		(1.524)	(2.567)	(7.859)	(5.783)	(3.365)	(5.620)
4 to <6 hours	68	6.132	11.456	50.853	34.882	40.029	21.294
		(1.337)	(2.634)	(6.890)	(5.692)	(3.515)	(5.920)
6 or more hours	29	6.103	10.586	48.931	35.276	39.862	21.379
		(1.235)	(3.168)	(8.298)	(5.619)	(4.095)	(5.766)
Age							
30-35	7	5.000	10.143	46.286	31.571	39.571	16.714
		(1.000)	(2.193)	(8.118)	(7.368)	(3.690)	(7.158)
36-40	106	5.406	11.104	47.094	32.670	39.453	18.434
		(1.632)	(2.711)	(9.062)	(6.374)	(3.420)	(5.752)
41-45	91	5.670	11.176	48.868	33.945	39.484	20.022
		(1.564)	(2.939)	(8.640)	(5.883)	(3.891)	(5.856)
46-50	42	5.810	10.929	50.143	35.381	40.071	22.500
		(1.742)	(3.564)	(8.478)	(5.738)	(3.432)	(5.270)
>50	19	6.263	11.526	50.316	36.053	38.895	22.256
		(1.759)	(2.836)	(8.374)	(6.720)	(4.771)	(6.141)
Branch of Service							
Air Force	244	5.578	11.020	48.234	33.582	39.541	19.635
		(1.615)	(2.881)	(8.626)	(6.207)	(3.568)	(5.918)
Army	2	6.000	13.000	50.500	34500	40.500	19.000
		(1.414)	(2.828)	(2.121)	(0.707)	(3.536)	(1.414)
Navy	0	N/A	N/A	N/A	N/A	N/A	N/A
Marines	10	5.600	10.800	48.700	34.800	37.600	21.700
		(2.271)	(4.077)	(13.483)	(7.525)	(6.022)	(7.150)
International	1	5.000	14.000	39.000	33.000	41.000	15.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Civilian	8	6.625	13.250	53.625	37.500	39.525	25.625
		(1.303)	(1.909)	(6.865)	(5.806)	(3.686)	(3.583)

Table 8 continued

Background	Sample size (N)	Perceived Effectivenes s $\bar{x}$ (S.D.)	Usability $\bar{x}$ (S.D.)	Attention $\bar{x}$ (S.D.)	Relevance $\bar{x}$ (S.D.)	Confidence $\bar{x}$ (S.D.)	Satisfaction $\bar{x}$ (S.D.)
Type of Service							
Active Duty	213	5.493	11.122	47.962	33.423	39.521	19.174
		(1.615)	(2.784)	(8.921)	(6.211)	(3.666)	(5.971)
Reserve	17	6.529	11.529	50.235	34.412	38.765	22.588
		(0.875)	(2.625)	(8.159)	(5.842)	(3.930)	(4.302)
Guard	21	5.667	9.762	49.238	35.048	40.048	21.905
		(2.129)	(4.242)	(8.390)	(5.937)	(3.263)	(5.787)
Civilian	14	6.214	12.357	51.500	36.000	39.714	24.143
		(1.369)	(2.469)	(7.763)	(7.211)	(4.480)	(4.834)
Occupation	182	5.337	10.711	45.542	32.036	39.205	17.675
Rated		(1.647)	(3.210)	(10.464)	(6.799)	(3.869)	(6.218)
Non-rated	265	5.736	11.286	49.698	34.533	39.670	20.874
		(1.614)	(2.772)	(7.585)	(5.803)	(3.600)	(5.585)
Wargame Experience							
Yes	211	5.597	11.066	47.929	33.630	39.573	19.479
		(1.599)	(2.904)	(8.954)	(6.231)	(3.634)	(5.885)
No	54	5.667	11.259	50.222	34.222	39.333	21.407
		(1.770)	(3.017)	(7.904)	(6.254)	(3.909)	(6.092)

Analysis of Research Question 1

Research question 1 asked:

Is the cultural simulation program perceived as effective?

The survey question that addressed this issue had a numerical range from one to eight. A mean rating above 4.5 indicated it was perceived as effective and a rating below 4.5 indicated it was not perceived as effective. A one-sample *T*-test was accomplished to investigate this question.

The assumptions for the *T*-test were the sample data were independent, there were no outliers, and the dependent variable was normally distributed (Lund & Lund, 2013). The data for the original simulation met those assumptions but the data for the VEST simulation did not meet the outlier assumption. Throughout this research, three standard deviations were used as the criteria for identifying outliers. The detailed assumption analysis is presented in Appendix C. The seven outliers in the VEST dataset had perceived effectiveness rated at the lowest possible rating of one. These represented 2.6% of the VEST surveys. All seven samples had open-ended comments that were negative about the simulation such as "It was a waste of time for someone with my experience," "couldn't understand a single word," and "it kept crashing." There was no reason to suspect data entry errors. For comparison, the original simulation had 96 ratings of 1 out of the 898 samples. These represented 10.7% of the original simulation surveys. However, the rating of 1 was not an outlier for the original simulation. The mean of the VEST ratings for perceived effectiveness was 5.611 so removing the lower ratings would slightly increase the mean, thus increasing the difference between the neutral rating of 4.5 and the VEST rating. Additionally, removing the outliers would decrease the variance and increase the risk of Type I error. Since the ratings appeared to be genuine and removing them would only increase the risk of Type I error, they were retained in the dataset.

The results of the *T*-test are presented in Table 9. Based on the *T*-test *p*-value of less than 0.001, the mean rating of 4.049 for the original simulation was statistically, significantly lower than the neutral value of 4.5, indicating the original simulation was not perceived as effective. The *T*-test for the VEST simulation also had a *p*-value of less than 0.001. The VEST mean rating of 5.611 was statistically significantly higher than the neutral 4.5 rating, indicating that VEST was perceived as effective.

Table 9

Perceived Effectiveness Ratings by Simulation Type

Type of simulation	Sample Size (*N*)	Mean	Std Dev	*T*-Test Significance
Original Simulation	898	4.049	1.82	0.000
New Simulation (VEST)	265	5.611	1.63	0.000

Although perceived effectiveness is not the same as actual effectiveness, it did provide insight into the student's perspective on the simulations. The lower rating for the original simulation and higher rating for the VEST simulation provided a clear indication that the perception was specific to each simulation rather than rating the general use of simulations for education, and the ratings indicated students could and did perceive educational simulations could be effective. The specific results for each simulation also provided a backdrop that could help in understanding the results of the remaining analysis.

Analysis of Research Questions 2, 3, and 4

Research questions 2, 3, and 4 were grouped together for analysis because they involved similar statistical tests to be performed. Research question 2 asked:

> Does prior video game experience impact perceived effectiveness, usability, attention, relevance, confidence, and satisfaction when using role playing simulations for PME?
>
> Research question 3 asked:
>
> Does gender impact perceived effectiveness, usability, attention, relevance, confidence and satisfaction when using role playing simulations for PME?

Research question 4 asked:

> Is there an interaction between gender and gaming experience on perceived effectiveness, usability, attention, relevance, confidence and satisfaction when using role-playing simulations for PME?

These were analyzed first with a factor analysis of Keller's (2010) IMMS categories, then with a test of assumptions, followed by MANOVA analyses of the original simulation and VEST datasets.

Factor Analysis

The first step in analyzing this data involved performing a factor analysis on the IMMS portion of the survey questions to confirm that a four-factor structure existed for the survey's motivation questions. Extraction was accomplished based on using eigenvalues greater than 1 and the varimax method was used for the rotation. Details are presented in Appendix D. Attention, relevance, and satisfaction all appeared to be loaded the highest on one component and confidence seemed to be loaded on a different, second component. It appeared that confidence was different than the other three but there was no indication within the data that attention, relevance, and satisfaction were clearly distinct from each other.

This did not negate the extensive research on Keller's (2010) IMMS but this particular case suggested a two factor model be used for analysis since there was no clear distinction in the variance for the variables of attention, relevance, and satisfaction. Those three variables were combined for the two-way MANOVA analysis.

Original simulation assumptions check. The first MANOVA analysis examined survey data from participants who completed the original simulation. The sample size was 898. Assumptions were checked to determine if the MANOVA test was appropriate. According to Lund and Lund (2013), there are two requirements and seven assumptions to check when using a MANOVA statistic; independent variables are categorical and dependent variables are continuous independent observations, adequate sample size, no univariate or multivariate outliers, multivariate normality, linear relationship, homogeneity of variance-covariance matrices, and no multicollinearity. Details of the assumption testing are provided in Appendix E and a summary of the results is presented in Table 10. Of note, there were 17 univariate outliers and three multivariate outliers.

Table 10

Results of Requirements and Assumptions Check

MANOVA Assumptions	Univariate	Multivariate
R1. DV measured as interval	✓	✓
R2. IV are categorical	✓	✓
A1. Independent Observations	✓	✓
A2. Adequate Sample Size	✓	✓
A3. No Outliers	17 identified	3 identified
A4. Normality	✓	✓
A5. Linearity	✓	✓
A6. Homogeneity of Variance	✓	✓
A7. Multicollinearity	✓	✓

Seventeen univariate outliers were identified using a boxplot and the outliers were all based on low ratings for the confidence variable. In total, there were 16 male outliers and one female outlier. The survey contained nine confidence questions. The Likert response value range for each question was from 1 to 5. Therefore, the net confidence value could range from 9 to 45. They all rated confidence at a value less than 20. Most of the written remarks for these outliers contained comments consistent with low confidence ratings including issues like the long time required to complete the simulation, the student not being sure how to get through the simulation, and the use of adjectives such as "hard," "frustrating," and "difficult." These appeared to be genuine data points. Simply removing these points carried risk of altering the results without knowing what effect the outliers had on those results. Kruskal (1960) recommended if outliers are present in the data, then an analysis should be completed with the outliers present and a second analysis should be completed with the outliers removed. If the results are similar, then there should be confidence in the results but if they are different, then the conclusions would be suspect.

Three multivariate outliers were identified by using the Mahalanobis distance method with a critical value of 18.47 based on a chi-square distribution and four degrees of freedom due to the four dependent variables (Lund & Lund, 2013). These were case numbers 253, 246, and 135. While close scrutiny of each outlier was necessary, they were multivariate outliers; they were outliers because of a combination of the dependent variables, which made them more difficult to assess than the univariate outliers. The demographics and numerical values for each of these multivariate outliers are listed in Table 11.

Table 11

Values for Multivariate Outliers

Sample Number	Mahalanobis Number	Gender	VGE	Perceived Effectiveness	Usability	Confidence	Combined SAR
253	29.4	Male	Experienced	8	8	28	50
246	20.7	Male	Experienced	5	1	13	49
135	19.1	Male	Inexperienced	1	1	9	29
			Possible Range	1 to 8	1 to 15	9 to 45	27 to 135
			Mean	4.05	8.42	33.5	81.6

Note. SAR = Satisfaction, attention, and relevance.

The furthest multivariate outlier was sample number 253 with a Mahalanobis number of 29.4, which was well above the 18.47 critical value cutoff. This sample was from a male in the experienced video gamer category and appeared to be an outlier because he rated the simulation the highest possible for perceived effectiveness but below average for all other variables. His wording on the open-ended questions centered on the simulation being cumbersome to load and often crashing. It is possible he thought the simulation was effective if he looked past the loading/crashing issues but perhaps those issues impacted his motivation. This sample could not be dismissed as a simple entry error.

The next furthest multivariate outlier was sample number 246 with a Mahalanobis number of 20.7. This sample was a male categorized as an experienced video gamer. He gave a higher than average rating for perceived effectiveness but gave the lowest possible rating for usability as well as low ratings for confidence and the combined SAR variable. His remarks mentioned the hints within the simulation were not helpful. It might be that he considered the simulation effective but difficult to use. Like the previous multivariate outlier, this did not appear to be a data entry issue.

The closest multivariate outlier to the critical value was sample number 135 with a Mahalanobis number of 19.1, which was also above the critical value of 18.47. This sample was also a male categorized as inexperienced with video games and had the lowest possible ratings for perceived effectiveness, usability, and confidence. It also had a very low rating for the combined SAR variable. For the open-ended question about what he liked the most about the simulation he answered "NOTHING" and for the open-ended question about what he liked the least, his answer included "This exercise was a huge source of frustration and had no relevance..." The ratings were consistent with these remarks so it did not appear to be a data entry error.

Even though three outliers out of 898 samples was a small percentage, the same solution identified for the univariate outliers was used. Two of the multivariate outlier surveys also contained univariate outliers so the total number surveys removed from the dataset for the second MANOVA test was 18. The 18 outlier surveys accounted for just 2% of the 898 samples.

Original simulation multivariate analysis of variance tests. Based on the results of the assumption testing, the statistical plan was modified to include a second iteration of the MANOVA to determine the impact of the outliers. A two-way MANOVA was accomplished for the dataset with the outliers present ($N = 898$) and a second MANOVA was accomplished with the outliers removed ($N = 880$) to test the effect of the outliers.

Based on the results of the comparison of the two MANOVA results, a stepwise discriminant analysis (SDA) was accomplished on the original dataset (N = 898) to determine which dependent variables were impacted by the independent variables.

Two-way multivariate analysis of variance with outliers present. A two-way MANOVA was run using the original dataset (N = 898) with the outliers present. As shown in Table 12, all variables had a higher mean for experienced video gamers compared to inexperienced video gamers. Across the four variables, there did not seem to be a trend with means based solely on gender.

Table 12

Descriptive Statistics for N = 898 Dataset

Gender	VGE	N	Perceived Effectiveness $\bar{x}$(s)	Usability $\bar{x}$(s)	Confidence $\bar{x}$(s)	Combined SAR $\bar{x}$(s)
Male	Inexperienced	446	3.888	8.027	32.760	80.360
			(1.793)	(3.435)	(5.845)	(21.937)
Male	Experienced	346	4.237	9.168	34.335	82.840
			(1.842)	(3.231)	(5.273)	(22.132)
Female	Inexperienced	55	3.964	7.491	32.945	82.330
			(1.856)	(3.706)	(5.349)	(19.568)
Female	Experienced	51	4.255	7.784	34.098	83.980
			(1.809)	(3.431)	(5.166)	(20.712

The MANOVA statistical test results are shown in Table 13. The Wilks-Lambda was used for all the MANOVAs. Both gender and VGE showed a statistically significant effect on the combined multivariate variable. However, the partial eta squared indicated the effect was small and was estimated to account for less than 2% of the variation. The interaction of gender and VGE did not indicate a statistically significant effect on the multivariate variable. Before accomplishing the SDA to determine which specific dependent variables were impacted by the independent variables, a second MANOVA was accomplished to determine if outliers affected the results.

Table 13

Two-Way Multivariate Analysis of Variance Statistics with Outliers

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
	.060	3502.659	4.000	891.000	.000	.940
Gender	.986	3.168	4.000	891.000	.013	.014
VGEy	.989	2.513	4.000	891.000	.040	.011
Gender VGE	.998	.410	4.000	891.000	.802	.002

N =898

Two-way multivariate analysis of variance statistics with outliers removed. The 18 outliers were removed from the original dataset changing the *N* to 880. A reexamination of the assumptions was accomplished. No additional outliers appeared and all the assumptions were met. The two-way MANOVA was run using the adjusted dataset (*N* = 880) with the outliers removed. This was accomplished to create a comparison for the effect of removing the outliers. As shown in Table 14, the means changed slightly except for the female experienced video gamers group which remained the same since it had no outliers in the original database. Confidence values increase in the other three groups. Males who were inexperienced with video games accounted for 14 of the 18 outliers and had a larger change in the mean. However, the same pattern persisted with experienced video gamers providing higher ratings for confidence. Standard deviations decreased as would be expected with the removal of outliers.

Table 14

Descriptive Statistics for N = 880 Dataset

Gender	VGE	*N*	Perceived Effectiveness $\bar{x}$(s)	Usability $\bar{x}$(s)	Confidence $\bar{x}$(s)	Combined SAR $\bar{x}$(s)
Male	Inexperienced	432	3.956	8.164	33.273	81.502
			(1.774)	(3.376)	(5.161)	(21.265)
Male	Experienced	343	4.233	9.216	34.464	83.146
			(1.830)	(3.191)	(5.079)	(21.914)
Female	Inexperienced	55	3.964	7.491	32.945	82.330
			(1.856)	(3.706)	(5.349)	(19.568)
Female	Experienced	50	4.260	7.860	34.440	83.900
			(1.827)	(3.423)	(4.599)	(20.914)

The MANOVA statistical results, shown in Table 15, were only slightly different than the previous MANOVA results and the effective results were the same. The *p*-values for gender and VGE both indicated a statistical significance on the combined multivariate variable, but partial eta squared results indicated the magnitude of the impact was small. Additionally, the interaction of VGE and gender did not appear to impact the dependent variables. These results were similar to the results of the original dataset, adding confidence to the previous MANOVA results.

Table 15

Two- Way Multivariate Analysis of Variance Statistics with Outliers Removed

Effect	Value	*F*	Hypothesis df	Error df	Sig.	Partial Eta Squared
	.050	4156.872	4.000	873.000	.000	.950
Gender	.986	2.988	4.000	873.000	.018	.014
VGE	.987	2.785	4.000	873.000	.026	.013
Gender VGE	.998	.438	4.000	873.000	.781	.002

N = 880

Stepwise Discriminant Analysis

To further investigate the effect and identify which dependent variables were most responsible for this effect, a stepwise discriminant analysis (SDA) was accomplished. The assumptions for the SDA were the same as the MANOVA so no additional assumption testing was necessary. The same approach to deal with the outliers utilized for the MANOVA was followed for the SDA. The analysis was accomplished with the outliers present (*N* = 898) and with the outliers removed (*N* = 880). The *p*-value for variables to enter into the function for the SDA was set at 0.05 and the *p*-value to remove the variables was set at 0.10.

The SDA was first performed using the VGE variable to provide insight into research question 2. The first step identified usability as a factor and the second step retained usability and added confidence. The two remaining variables did not have a low enough *p*-value to be added to the function. The Wilks-Lambda significance for the resulting function was less than 0.001 so it was reasonable to assume the function explained the variation. The canonical correlation was 0.163. The square of the canonical correlation was 0.0266, indicating the two variables accounted for an estimated 2.66% of variation. Although the results were statistically significant, the total effect was small.

To check if the outliers impacted the results, the SDA was accomplished with the outliers removed (*N* = 880). The first step identified usability. The SDA stopped after one step because the *p*-value for confidence was 0.092, which was too high to enter. This compared to a *p*-value for confidence of .041 after the first step in the previous analysis. The Wilks-Lambda for the function had a significance of less than 0.001. There appeared to be solid indications that the usability variable was affected by VGE. The difference between the two SDA results regarding the confidence variable was understandable given the 17 univariate outliers that were removed were all due to low scores in confidence. Although confidence outliers appeared to be valid, the link between video game experience and the confidence variable identified by the first SDA analysis was questionable due to violating the SDA assumption regarding outliers.

Gender was then analyzed using the SDA to further investigate research question 3. Using the database with the outliers present (*N* = 898), step one resulted in usability being added to the function and step two resulted in retaining usability and adding the combined SAR variable. The iterations stopped after two steps because neither of the remaining variables had a low enough *p*-value to enter. The Wilks-Lambda significance for the resulting function was .005 so it was reasonable to assume the function explained the variation. Running the analysis with the outliers removed ($N = 880$) resulted in only step one being accomplished. The usability variable was identified in the first step. The analysis stopped after the first step because the *p*-value for the combined SAR was 0.087, which was too high to enter into the function. The combined SAR *p*-value for the second step with outliers present was 0.036. The significance of the function was less than 0.001. There appeared to be solid indications that the usability variable was affected by gender. However, the link between gender and the combined SAR variable was questionable. The canonical correlation for the $N = 898$ dataset was .109 and for the $N = 880$ dataset, it was .093. Those numbers equated to approximately 1.19% and 0.86%, respectively, so there was consistency for the estimated effect size. Similar to VGE, gender showed a statistically significant effect but the effect size seemed to be very small.

Research question 4 dealt with the interaction of gender and VGE. The interaction was not investigated with a SDA because the MANOVA did not find a statistically significant result.

Visual Expeditionary Skills Training Analysis

In this section, examination of the second, third, and fourth research questions is presented similar to the previous section but using the VEST dataset. First, the VEST dataset is described and then the assumptions test and the two-way MANOVA are presented and compared with the previous MANOVA results.

Visual expeditionary skills training dataset demographics. The VEST dataset was smaller than the original simulation. It only contained 265 surveys, 35 of which were from females. Females accounted for approximately 13% of the VEST respondents, which was a slightly higher percentage than the original simulation. However, due to the low total number of females in the VEST dataset, the median split points to divide the subjects into experienced and inexperienced categories were recalculated to ensure the split points were still appropriate. The median split points remained the same with 5 and above for males and 4 and above for females for the experienced video gamer category. This resulted in 148 inexperienced and 117 experienced video gamers. As shown in Table 16, there were only 14 surveys for females who were experienced video gamers.

Table 16

Demographics for Visual Expeditionary Skills Training Database

VGE	Female	Male	Total
Inexperienced	21	127	148
Experienced	14	103	117
Total	35	230	265

Visual expeditionary skills training assumption testing. The MANOVA assumptions were examined with the VEST dataset. The details of the process are presented in Appendix F and a summary of the results is presented in Table 17. The VEST dataset contained outliers and there were issues with normality for the confidence variable. All other assumptions were met.

Table 17

Results of Requirement/Assumption Check for Visual Expeditionary Skills Training

MANOVA Requirements/Assumptions	For Each DV	Within Groups
R1. DV measured as interval	✓	✓
R2. IV are categorical	✓	✓
A1. Independent Observations	✓	✓
A2. Adequate Sample Size	✓	✓
A3. No Outliers	25 Outliers	3 Outliers
A4. Normality	Confidence	Confidence
A5. Linearity	✓	✓
A6. Homogeneity of Variance	✓	✓
A7. Multicollinearity	✓	✓

N = 265

A total of 17 surveys contained univariate outliers. Some surveys had outliers for two or more variables so a total of 25 ratings within the 17 surveys were outliers. There was no indication these outliers resulted from data entry issues. All univariate outliers were the results of low ratings. Perceived Effectiveness had seven outliers. Usability had seven outliers. Confidence had eight outliers and the combined SAR variable had three outliers. Comments generally focused on two issues. First, the simulation ran extremely slow due to bandwidth issues and second, some people with prior experience in dealing with cultures felt they should not have to participate in the simulation.

The VEST dataset contained a higher percentage of outliers than the original simulation dataset. Seventeen of the 898 original simulation surveys contained outliers, which was approximately 1.9% of the surveys. Seventeen of the 265 VEST surveys contained outliers, which equated to approximately 6.4% of the VEST surveys. Additionally, outliers were present in several of the variables for this VEST dataset compared to the original simulation dataset, which only had outliers for Confidence. This was consistent with the higher ratings and lower standard deviations for the VEST data. There was a tighter grouping of ratings toward the upper part of the scale compared to the original simulation, which resulted in outliers for scores that would not have been outliers in the original simulation.

To check for multivariate outliers, a Mahalanobis distance was computed for each sample. Three samples were above the critical value of 18.47, indicating three multivariate outliers as shown in Table 18. All three outliers were also univariate outliers. Therefore, the comments dealt with the two issues of download speed and previous expertise with culture. To deal with the 17 samples containing outliers if a statistically significant result was identified with the original $N = 265$ dataset, a second dataset with outliers removed would be analyzed to help determine the impact of the outliers. The comparative dataset with the outliers removed had a sample size of 248.

The dependent variables were checked for normality. The confidence variable showed more of a departure from normality than the other variables. The departure from normality for confidence should be considered with the results of the MANOVA.

Table 18

Values for Multivariate Outliers for Visual Expeditionary Skills Training

Sample Number	Mahalanobis Number	Gender	VGE	Perceived Effectiveness	Usability	Confidence	Combined SAR
184	28.4	Male	Experienced	6	5	31	37
98	23.5	Male	Experienced	1	1	24	42
75	19.1	Male	Inexperienced	1	1	37	76
			Possible Range	1 to 8	1 to 15	9 to 45	27 to 135
			Mean for VEST	5.6	11.1	39.5	102.0

N = 265

Visual expeditionary skills training two-way multivariate analysis of variance with outliers present. The two-way MANOVA was run using the VEST (*N*= 265) dataset. This dataset had outliers present. The groups for this dataset were much smaller than the dataset for the original simulation. Females who were experienced video gamers had a group size of just 14. The descriptive statistics shown in Table 19 display a pattern for females with less video game experience having the highest mean for every dependent variable. This trend in the data contrasted sharply to the original simulation trend in which females who were inexperienced video gamers had the lowest means for perceived effectiveness and usability and the second lowest for confidence and the combined SAR variable. The small group size for females might be the source of this inconsistency.

Table 19

Descriptive Statistics for the Visual Expeditionary Skills Training Dataset with Outliers Present

Gender	VGE	*N*	Perceived Effectiveness $\bar{x}$(s)	Usability $\bar{x}$(s)	Confidence $\bar{x}$(s)	SAR $\bar{x}$(s)
Male	Inexperienced	127	5.53	11.01	39.63	102.07
			(1.65)	(3.07)	(3.62)	(18.18)
Male	Experienced	103	5.65	11.18	39.26	100.20
			(1.65)	(2.66)	(3.84)	(20.96)
Female	Inexperienced	21	6.00	11.81	40.52	109.81
			(1.34)	(3.25)	(3.70)	(15.97)
Female	Experienced	14	5.50	10.36	39.00	103.21
			(1.62)	(2.95)	(3.09)	(18.04)

N = 265

To more fully examine the differences in the means, a MANOVA was accomplished. The statistical results, shown in Table 20, indicated no statistically significant effect due to VGE, gender, or the interaction of those two variables. Since no statistically significant effect was identified, a confirmatory MANOVA with the outliers removed was not conducted.

Table 20

Two-Way Multivariate Analysis of Variance Statistics with Outliers

Effect	Value	*F*	Hypothesis df	Error df	Sig.	Partial Eta Squared
	.018	3527.562	4.000	258.000	.000	.982
Gender	.985	.986	4.000	258.000	.415	.015
VGE	.991	.594	4.000	258.000	.667	.009
Gender VGE	.988	.811	4.000	258.000	.519	.012

N = 265

Research question 2 dealing with VGE had mixed results between the two simulations. The original simulation analysis found a statistically significant impact of VGE on usability. Those respondents who had more video game experience rated the simulation as more usable. This was the expected outcome but the estimated effect size was small, measuring just under 3%. Although the SDA also found a statistically significant result for confidence, the outliers made this finding questionable. The VEST analysis did not find a statistically significant effect for video game experience but likely reflected the very small group sizes and the small effect size.

Research question 3 dealing with gender also had mixed results between the two simulations. The original simulation dataset analysis identified a statistically significant impact of gender on usability. Males generally rated the simulation as more usable. The expectation was there would be no statistically significant difference based on gender. The estimated effect size was very small, measuring around 1%. The original simulation dataset also identified a statistically significant effect on the combined SAR variable.

However, the combined SAR variable was only identified when the outliers were present in the dataset, making that result questionable. The VEST analysis with the much smaller sample size did not find a statistically significant effect due to gender.

Research question 4 dealing with the interaction of the two independent variables was consistent. Neither analysis found a statistically significant effect due to the interaction of gender and VGE. While the original simulation analysis found usability was impacted by both gender and VGE, those impacts were parallel. Experienced gamers for both genders had higher means and males for both VGE categories had higher means.

Analysis of Research Question 5

The final research question asked,

What aspects of the simulation did the students find valuable or problematic?

To assess this question, two open-end questions were included in the survey. Out of the 898 surveys for the original simulation, 709 surveys provided meaningful feedback for question 52: “What did you like the most about the simulation?" Seven hundred sixty-two surveys provided meaningful feedback for question 53: "What did you like the least about the simulation?" One comment from the discarded incomplete surveys was included in this analysis. The responses from the VEST simulation were analyzed separately and those results are presented after the original simulation comment analysis.

Open-ended comments by their very nature can cover a wide variety of topics. To better understand the meaning of the comments as a whole, they were sorted by themes and then subareas were developed within the themes. Some survey comments were a single sentence or phrase, but many comments covered multiple themes and subareas. As a result, the total number of the responses by theme was greater than the total number of surveys. The basic themes, number of responses, and percentage of the total number of surveys are presented in Table 21. The total number of surveys used to compute the percentage of surveys for the original simulation was 899, reflecting the addition of the comment from the one incomplete survey. Although the terms *favorable* and *unfavorable* are used in this discussion, it does not imply that every comment in the negative column identified a problem. Some were merely suggestions on things they would like to see such as additional scenarios. Likewise, every comment in the positive column was not necessarily praising the simulation. Comments like "Completing it," which is discussed later, might actually have had a negative connotation. Additionally, while the percentage of respondents was considered, it must be acknowledged that a large number of comments did not necessarily identify a valuable insight and a small number of comments in a given area might provide valuable insight. However, the numbers did provide an additional basis to evaluate and garner insights from the comments. Finally, it must be acknowledged the comments included phrases and sentences that were less than clear in some cases. These had to be subjectively evaluated to determine what message the respondent was trying to convey. This is simply the nature of working with open-ended questions.

Table 21

Original Simulation Macro-View of Themes/Comments

Liked "Most"	# of Responses	% of Surveys	Liked "Least"	# of Responses	% of Surveys
Preparation and Completion	10	1.1	Preparation and Completion	138	15.4
Simulation Game Play	194	21.6	Simulation Game Play	660	73.5
Simulation Technical Quality	77	8.6	Simulation Technical Quality	97	10.7
Content	293	32.6	Content	160	17.8
Overall	384	42.8	Overall	77	8.6
Blank or "Nothing" comments	185	20.6	Blank or “Nothing" comments	136	15.1

As mentioned, the comments were categorized into five major themes. The first theme was Preparation and Completion. This theme covered aspects of preparing to use the simulation and getting credit for completing the simulation. These issues are an important part of an educational simulation but do not directly relate to the learning environment within the simulation. The second theme dealt with Simulation Game Play that looked at the ability to operate the simulation, how well the simulation design worked, which in this case incorporated multiple branches, and the impact of the hints and feedback within the simulation to facilitate learning and guide the player. Although feedback could be considered part of content, it was included within the Simulation Game Play theme because quality, accessibility, and timing of hints and feedback impacted how the student experienced the simulation regardless of content. The third theme dealt with Simulation Technical Quality including the graphics, sounds, and how well it operated. These were issues that dealt with the current technology of the simulation and did not fit into the simulation game play theme that dealt with the design or underlying instructional method. The fourth theme covered Content. This included overall assessments of the content as well as specific remarks about the scenarios and characters. While there was an element of Simulation Game Play within scenarios and characters, the subareas including realism and importance were most directly tied to the underlying content of the subject material so they were included within this theme. The final theme was Overall comments about the simulation experience. These included perspectives on the simulation and what the participant felt he/she learned. Each of these five themes was further broken down into subareas, which are discussed in the following paragraphs. The first four themes were more tactical in nature, focusing on the specific simulation. However, those comments still provided insights that could help in creating educational simulations. The fifth theme provided a more strategic look. While comments in this theme were influenced by this particular simulation, the comments provided insights into the general value of using simulations for education and how they are perceived.

Among the five themes, Content and Overall received the largest percentages of favorable comments. The largest area for unfavorable comments was Simulation Game Play. The least commented on areas were Preparation and Completion and Simulation Technical Quality.

Theme 1: Preparation and Completion

The simulation required users to download and install the simulation to their work or home computer to operate the simulation. When the simulation was completed, a code was generated by the simulation for the participant to load onto the Air War College website as proof of completion. Access and proof of completion are important aspects of educational simulations but are distinct from the actual operation of the simulation, which is where the learning takes place. Additionally, the directions for which keys to press to navigate within the simulation do not generally contribute to the learning but can be critical to enable learning from the simulation experience. The survey questions asked what the participant liked most or least about the simulation. Because these subareas dealt with preparing to use the simulation and proving simulation completion rather than pertaining to the actual simulation operation, content, or experience, these subareas might have been underreported among both the favorable and unfavorable comments due to the wording of the question. As Table 22 shows, there were 138 unfavorable responses in this category but only 10 favorable comments. This disparity in types of responses was reasonable since participants who had no problems in this area would be unlikely to choose to comment about it for the reasons already noted. However, those who had a serious problem in this area might have felt frustrated with the process of resolving the issue and as a result, they might have been very motivated to ensure the problem was captured on the survey. Overall, 15.4% of surveys noted problems in this area.

Table 22

Original Simulation: Preparation and Completion Comments

Liked Most	# of Responses	% of Surveys	Liked Least	# of Responses	% of Surveys
Total Preparation and Completion	10	1.1	Total Preparation and Completion	138	15.4
Preparation			Preparation		
No Tech problems	1	0.1	Problems downloading	11	1.2
Directions were good	8	0.9	Problems installing	19	2.1
Tutorial was helpful	1	0.1	Problems with Initial Directions	45	5.0
			Problems with direction to save	10	1.1
			Completion		
			Didn't know when done	33	3.7
			Issues with completion Codes	20	2.2

Examining the unfavorable subareas identified issues with downloading and installing that could have many causes. Comments were generally similar to these: "I had a lot of trouble loading it and had to use 3 different computers before I got it to work correctly," "My Internet Explorer anti-virus software/firewall locked me out of AUSIS website and it took me about 1 hour to find a way in and download the sim software," and "Didn't work on my Mac." It is worth noting that the simulation program was written for the Windows environment and could not be used on a Macintosh computer. This left some students frustrated and several noted the issue in their comments. The decision to make the program Windows-based was made with the idea that those participants who did not own a Windows-based computer at home could download and install it on military work computers that were Windows-based. However, even though the program was initially approved to install on military computers, every time there was a service pack update or transition to a new operating system (Windows XP to Windows 7), the program had to be recertified, which could be a lengthy process. During those times when the program was awaiting recertification, students needing to use their work computer to accomplish the simulation had to wait. Finally, installing programs might have been difficult for participants who were not computer savvy. The approach of creating a downloadable program had benefits such as (a) not having to be connected to the Internet to operate the simulation and (b) being less dependent on bandwidth since downloading only required the user to connect to the Internet one time. However, as the comments revealed, using an installation-based program rather than a web-based or CD-based program could potentially have created significant issues in preparing to use the simulation. The other simulation program (VEST) was web-based, and the results from those comments are presented later along with a discussion of the merits of both approaches.

The subarea of directions received comments from 6.1% of the surveys. Representative favorable comments were "The instructions were well documented in the associated lesson and the simulation moved along quickly" and "Well organized. Instructions were long, but very useful and complete." Some representative negative comments were "The directions were a little confusing," "Instructions were a bit cumbersome," and "The instructions were poorly written. I spent extra frustrating hours on the simulation because of the instructions." It is worth noting that in the actual comments, one of the students who thought the directions were good mentioned, "If you read the instructions, using the simulation was easier than I expected." Many students who mentioned problems with the directions or with the mechanics of saving files mentioned they skimmed the instructions such as the student who noted, "I must admit, I did not read the complete instructions..." and one who said, "The time stress from my full time duties and family led me to skim the directions and waste time repeating scenarios." The issue of directions might have been less about what the directions said and more about how the information was presented. Among the files that were downloaded for the simulation program, there was a 45 page player's guide in portable document file (PDF) format that could only be accessed from outside of the simulation environment. This file covered the mechanics of playing the simulation, the behavioral dimensions of culture, as well as scoring and success criteria. Thoroughly reading these directions was probably very useful as some of the participants noted. Additionally, a tutorial could be accessed from the main menu that guided the user through each of the navigation keys within the simulation environment. The tutorial also covered how to save progress within the simulation. The tutorial was not played by default. The player had to notice the option and specifically select it. Although the tutorial was set in the simulation environment, it focused exclusively on teaching the keys to control the characters and did not present any cultural issues. Many video games today are designed to be used without a manual and the tutorial is embedded within the first part of the game, instructing the user on how to play the game while providing a sense of making progress in the game. The poor ratings on directions might have been due to the separate 45 page player's guide being too long, not accessible from within the simulation as well as the tutorial being accessed only as a separate module. These issues might have led to many users not becoming familiar with the directions before using the simulation. This in turn might have had a negative impact on their experience as many comments mentioned frustration with the process of trying to save their progress and with uncertainty in how to accomplish some of the less intuitive actions such talking to a second character on the screen. Given the trend in commercial games not to include manuals and embedding the information within the first part of the game, serious consideration should be given to using a similar approach for educational simulations.

The final unfavorable subarea for the Preparation and Completion theme was completion, which involved two issues. The first was how the student was informed of satisfactorily completing the simulation program and the second issue dealt with generation and submission of the completion code itself to the Air War College. There were no positive comments in this area but as mentioned earlier, the lack of positive comments was reasonable since these were open-ended questions and both these areas were external to playing the simulation.

The simulation did not require students to complete every scenario and the point at which the student had completed enough of the simulation to receive credit varied depending on the choices the student made during the simulation. The student could satisfactorily complete the simulation with just four scenarios or if other choices were made during the program, up to three additional scenarios might be required. This seemed to have confused some students who unknowingly continued with the other scenarios after meeting the simulation requirements. Many students made comments similar to this one: "I had done enough to complete the exercise, early, but kept going for two hours, not knowing that I met the mins." The 45-page player’s guide contained information on identifying the completion point. When a satisfactory level of accomplishment had been reached during the simulation, (a) the indicator lights changed, (b) feedback included wording intended to convey the student had satisfied the simulation requirements, and (c) the Send Data button became active. Unfortunately, many students failed to spot these subtle indications and continued with the simulation. Student action that contributed to this problem included (a) not reading the directions, (b) not noticing or knowing how to interpret the indicator lights, and (c) not reading or not understanding the implication of the line in the feedback about satisfactory performance. However, given that the simulation was designed to have different completion points, consideration should be given to adjusting the graphical user interface to make it very clear to the student when the completion point has been reached. This could be as simple as a pop-up window that congratulated the student on achieving a satisfactory level of accomplishment in the simulation. The subtle nature of the completion indicators in the simulation created unnecessary frustration for students. As mentioned in the discussion on directions, designing the simulation in a way that does not require a written operating manual to get through the simulation might be a better approach.

The second part to this subarea was generating and submitting the completion code for credit. A representative comment was "Getting starting and finishing (putting code in AWC system) was very confusing...the instructions were spread out in different places...need to put a complete simulation guide on the web in one place." Generation and submission of the completion code was necessary but did not teach the student anything about culture so problems in this area only detracted from learning. To generate the code, the student entered his or her student number into the simulation. A coded string of numbers was then generated that, when decrypted by Air War College, provided the student's number and the choices they made during their successful completion of the simulation. This process ensured each student submitted a unique code upon completion of the simulation. However, the submission was not automatically accomplished by the program.

Additionally, as students noted in their comments, the procedures for uploading the code to Air War College were listed on the AUSIS website but were not provided in the written directions for the simulation or embedded within the simulation itself. The player's guide noted that upon reaching an acceptable level of performance by the end of the fourth scenario, the Send Data option would be enabled for the student to report completion of the simulation but no additional details were provided. Although the percentage of students reporting problems with this was low (2.2%), the actual number experiencing problems might have been higher. Consideration should be given to automating the process or better embedding the procedures within the simulation.

Theme 2: Game Play

The Game Play category contained 854 total comments, which was the largest number of comments from students for any theme--21.6% of students provided positive comments and 73.5% provided negative comments. These comments were divided into three subareas for analysis and the percentages for each of the areas are provided in Table 23. The first subarea was Navigation, which dealt with the physical process of moving the avatar around in the simulation, accessing various items, and saving progress. Problems with the instructions for saving progress were covered in the previous theme. This theme covered the actual process of saving progress within the simulation. The second subarea was Design, which included comments that dealt with the specific way the simulation was designed to teach with its multiple branches or paths to complete the program. The final subarea was Hints/Feedback, which dealt with prompting by the simulation program to help students understand what they should do if they became stuck and how they did upon completion of the scenario or simulation.

Table 23

Original Simulation: Game Play Comments

Liked Most	# of Responses	% of Surveys	Liked Least	# of Responses	% of Surveys
Total Game Play	194	21.6	Total Game Play	660	73.5
Navigation			Navigation		
Easy to navigate	14	1.6	Difficult to save progress	16	1.8
			Save points caused repetition	57	6.3
			Difficult to navigate	66	7.3
Design			Design		
How choices affect other choice	14	1.6	Too limited in choices	93	10.3
Multiple paths to success--not scripted	16	1.8	Choices unrealistic	21	2.3
Ability to save/restart/explore	50	5.6	Repetition	14	1.6
Interactive	63	7.0	Frustration: stuck in loop/too long	91	10.1
Quick/easy to get through	17	1.7	No acceptable path to recover	58	6.5
			Narrow path forced trial and error	131	14.6
Feedback/Hints			Feedback/Hints		
Character actions were interesting	1	0.1	Characters stopped talking	23	2.6
Hints were good	4	0.4	Need hints	43	4.8
Feedback was instantaneous	5	0.6	Need better feedback	47	5.2
Feedback was good	10	1.1			

Comments about navigation were short. Some simply stated it was "easy to navigate" or it was "difficult to navigate" without providing additional detail. However, many of the 66 negative comments about navigation specifically noted the "clunky" process of getting around and the navigation was not intuitive.

It should be considered that while many of the references to navigation clearly referred to the "physical" process of maneuvering the avatar to interact with the various characters, some of the comments might have referred to making the correct decisions during those interactions to "navigate" successfully through the scenario. The interaction to navigate through the simulation issue is covered later in the Design subarea. The actual "physical" maneuvering of the avatar was fairly basic--using arrow keys to move. However, commands to have the avatar sit or give objects involved keystrokes that were not as intuitive and talking to a second character within the scene involved mouse actions. While the keyboard commands were all contained in the written directions and were accessible by clicking on the "?" button on the screen at any time during the simulation, forcing the user to play through the tutorial could have helped solidify their understanding of how to navigate.

Some navigation functions were rarely used. Navigating to "talk" to a second character within a scene was only necessary for one specific scene that occurred in one of the later scenarios. The survey comments indicated that by the time the later scenario was reached, many students had forgotten the option to talk to a secondary character was even possible. The process might have been more intuitive if talking to the secondary character had been one of the choices provided when interacting with the primary character by presenting it within the normal course of navigation rather than relying on a rarely used mouse action. As mentioned earlier, serious consideration should be given to incorporating the operating information within the normal progression of playing the simulation. This would allow specific, rarely used navigation actions to be presented "just-in-time," allowing users to focus on learning the content rather than learning and memorizing how to navigate through the various aspects of the simulation.

The other issue mentioned within the navigation subarea dealt with how the simulation saved a student's progress. Positive comments for saving centered on liking the ability to save so those comments are covered under the Design subarea. Negative comments dealt with the process of saving. The following comment captured many of the concerns expressed on this issue: "I could not navigate the simulation properly in order to save my work and come back to it later. I could not immediately repeat the scenario that I had trouble with without starting the entire simulation over. This was frustrating."

The Save issue had two key aspects--saving the file and resetting the scene. First, the process of saving the file required the student to type in a file name. However, the place to type the file name was not presented with a typical "Save file as:" in front of the input field and the field itself did not stand out from the background like most input fields. Additionally, the program did not have a default file name for saved files so if users did not realize they had to type in a file name and instead just hit "save" without a file name, their progress was not saved and there was no warning screen identifying the failure to save progress. This resulted in many students being forced to go back to the beginning of the simulation. Based on the comments, some students never discovered how to properly save their progress, resulting in a great deal of frustration and a lot of extra time to complete the program. If properly saved, then the student could reload the saved file and begin at the start of the current scenario rather than having to start the entire simulation over again at the beginning the first scenario.

The second aspect was resetting the scene. The simulation was set up with seven scenarios. During each scenario, the student had a task that needed to be completed to further the humanitarian relief mission. The scenario started at the command tent where the student had to choose which characters to visit and the sequence of those visits. Because the sequence of visits was important, the Save function was set up to restart the entire scenario, allowing the student to recover from a poor choice on who was first visited. Information from one character could bring crucial insight into dealing with another character. Each visit was a "scene." The number of scenes in a scenario varied depending on the order of the visits and what transpired during the discussions. A Scene Reset function allowed students to recover from a poor choice during a discussion with a specific character without having to go back to the start of the scenario. Some students knew how to save their progress but either did not know either how to reset the scene or did not know it was even possible to reset the scene. Comments described frustration with the Save function because it returned them to the start of that scenario but they really wanted the ability to go back one decision point. The Scene Reset function provided that capability. However, utilizing the Scene Reset function was not intuitive. It was accessed by hitting the "Esc" key on the keyboard to get to the main menu and then selecting Reset Scene. This option did not appear on the main menu when initially accessing the simulation so unless the user knew it was there (either from the directions or noticing it when exiting or saving the program), he/she could complete the entire simulation without becoming aware of its existence.

The first aspect regarding the process of saving progress could have been resolved by using a default file name ensuring that selecting Save would always create a saved file. The second aspect regarding resetting the scene could have been improved by making a scene reset button available on the screen while playing the scenario rather than forcing the user to leave the playing portion of the simulation to access the main menu screen. These changes could potentially have prevented or attenuated some of the frustration students expressed regarding the Design, which is covered in the next section. With 8.1% of students choosing to discuss problems with saving their progress, it was clearly an area of concern.

The design used for this simulation was a multiple branch system. Beyond choosing who to visit and in what order, students also had to decide what to discuss and how to discuss it. For example, when visiting the airfield manager, the student was presented with four choices on what to say first to the character. This could vary from exchanging pleasantries to demanding action. The choice made had an immediate response affecting both the subsequent behavior of the characters and the available options within that scene. The choice could also have a long-term effect manifested in subsequent scenarios. As a result, the path students took to complete the simulation varied based on the decisions they made. There were multiple paths to successfully completing the simulation, but the full impact of bad decisions was not necessarily or immediately obvious. A poor decision might manifest a problem in the next scenario or even a couple of scenarios later. There were paths that were dead-ends due to a series of poor decisions. Students were then required to go back to an earlier scenario to implement a different decision. The simulation provided general guidance on problem areas but did not identify which choice the student needed to change. This was intended to force students to think about the subject area identified and determine on their own what alternate decisions or interaction changes were needed to fix the problem.

The Design subarea accounted for most of the positive comments for the Game Play theme. On 9% of the surveys, favorable comments were made regarding the multiple branch design. Respondents liked the inter-related decisions, the multiple paths to success, and using the Save capability to explore various decisions. Additionally, 7% of surveys gave favorable comments regarding the interaction created by the design. Some representative positive comments were "The interconnectivity of each level - especially since there were real impacts from different actions at each level," "The multiple pathways and possibilities that existed within the simulation," "I liked the option of trying new approaches to see the impact and then being able to reset the scene," and "The interaction between with the characters."

The Design subarea not only accounted for most of the negative comments for the Game Play theme but also accounted for more comments than any other subarea--positive or negative. Several aspects of the design were identified as problematic. There were 12.6% of respondents who did not like the limited/unrealistic choices available. Many mentioned none of the choices were ones they would pick. Another 14.6% of respondents felt they were forced to find a narrow path to completion using a trial and error method. Additionally, 11.7% of surveys mentioned frustration with being stuck in a loop, how long it took, or the repetition the simulation caused was a problem. Finally, 6.5% of the students disliked that there was no acceptable path forward to recover from a "dead end" path. Instead, students had to go back to a previous scenario to change one or more decisions in order to reach a successful conclusion. Some representative negative comments were "The choices I had for the dialogue. There was very rarely a choice that I would have chosen," "Though advertised differently there seemed to be too narrow a way to find success in the simulation," "Some scenarios came down to trial and error. It was frustrating to have to re-do some of the scenarios several times," and "The solutions were frustrating and it seemed like you could get caught in endless loops without a way out."

The sheer number of favorable and unfavorable comments in this area merited discussion. The design was clearly central to the simulation. Some of the unfavorable comments might have stemmed from students who did not understand how to save their progress. Having a more user-friendly Save function might have alleviated some of these negative experiences. Interestingly, the Trial and Error negative comments and the Explore other paths positive comments were two opposing viewpoints on the same capability. The difference was the perception of "having" to go back rather than "getting" to go back, which perhaps stemmed from the initial decisions being successful instead of feeling like their selections led to a "dead end." The comments about no acceptable path to recovery might be the key to resolving this issue in a more favorable way. Instead of forcing students to go back multiple scenarios to recover from incorrect or less desirable decisions, creating a successful path forward that included extra interactions with characters to provide additional opportunities to demonstrate they have learned how to successfully deal with specific topic might resolve the negative into a positive.

The final area for the Game Play theme was Feedback and Hints. The simulation provided some hints through the course of game play and provided feedback at the end of each scenario. The feedback was designed to be conceptual in nature. Rather than telling the student to change a certain decision when dealing with the airfield manager, the feedback discussed perspectives from other cultural viewpoints, allowing the student to analyze how they might alter one or more previous decisions to achieve the desired result. The intent was to improve cultural understanding by having the student reflect on the feedback while interpreting the character's actions to better understand the motivations and perspectives of the different characters. The process of determining what needed to change provided opportunities to learn and grow rather than merely giving feedback identifying a specific problem and providing an "approved" solution. This design was intended to make the student use critical thinking skills and deter them from randomly selecting a path, reading the feedback on which decisions to change and then redoing the simulation successfully possibly without really thinking about why a particular decision might be a poor approach. This method certainly had its risks. If the student did not understand the feedback, then he/she could have resorted to the trial and error method out of desperation or to save time and effort. Many survey responders either mentioned this directly or implied it in their comments. Two representative comments were " it was just trial and error...I don't feel I got much out of it" and "it was merely trial and error to see what episodic steps were required to pass each level." The way in which trial and error was noted in the comments made it clear they were referring to selecting choice A, and then if that did not work, resorting to choice B (or C, or D) rather than analyzing the choices with respect to desired outcome and determining a better path to achieve the mission objective. Using the trial and error process negated some or all of the learning value of the simulation. Additionally, some comments mentioned that students did not read the feedback. Failure to read the feedback might have been due to lack of time or lack of interest. In either case, choosing not to read the feedback would make the simulation far more difficult to complete without resorting to trial and error and would likely have a very negative impact on the simulation experience.

The Debriefing screen at the end of the fourth scenario also provided feedback, which was given in text format on the left side of the screen, and required using a drop-down menu to access each scenario's feedback. The feedback was general in nature. On the right side of the screen, the student could select and listen to several short audio lectures on culture. Potentially the feedback could have been improved by adding a video supplementing or replacing the text feedback. This might have improved the students' attention and understanding of the feedback. The feedback could present a different application of the same concept as an example to help the student better understand the concept.

Theme 3: Simulation Quality

Simulation Quality was the most balanced theme area in terms of comments and also had the least total number of comments. There were 8.6% of surveys that commented in a positive manner and 10.8% provided unfavorable comments as indicated in Table 24. In particular, the graphics and graphical user interface comments were surprisingly balanced with 51 positive comments and 49 negative comments. Examples of the wide range of comments dealing with graphics were "Good Graphics worked well," "Decent graphics," "Archaic graphics," "I felt like I was playing a 1980s video game," and "The graphics were awful and detracted from the simulation."

Additionally, 17 students praised the simulation quality with comments such as "Quality of the graphics was superb" and "Quality of the video was impressive."

However, 42 students had negative comments about the simulation quality. Comments included "Glitchy interface on occasion," "The horrible, buggy interface," " It was buggy and did not work well," and "It took too long to navigate between scenes because of the loading time."

Table 24

Original Simulation: Simulation Quality Comments

Liked Most	# of Responses	% of Surveys	Liked Least	# of Responses	% of Surveys
Total Simulation Quality	77	8.6	Total Simulation Quality	97	10.8
Great/Good GUI/Graphics	51	5.7	Antiquated GUI/Graphics	49	5.5
Great/Good Voice/sounds	9	1.9	Errors	2	0.2
High Quality	17	1.0	Poor script	2	0.2
			Voice/sound issues	2	0.2
			Glitchy, Froze-up, Buggy	29	3.2
			Long loading times	13	1.4

The comments regarding long loading times when switching between scenes could be due to problems external to the simulation such as using a less powerful computer rather than a problem with the simulation program itself. Overall, the percentage of students commenting on the loading time problem was low. There were additional comments on voice, sounds, errors, and script but they were very low numbers, indicating it was not a significant issue for most students.

Theme 4: Content

There were approximately twice as many favorable comments about content than there were unfavorable comments. As shown in Table 25, 32.6% of surveys contained comments identifying things they liked about content and 17.8% of surveys had comments on things they liked the least about content. While many of the subareas had very small percentages, they helped provide a sense of the variety of responses.

Table 25

Original Simulation: Content Comments

Liked Most	# of Responses	% of Surveys	Liked Least	# of Responses	% of Surveys
Total Content	293	32.6	Total Content	160	17.8
Overall			Overall		
Good support materials	6	0.7	Too simplistic/Too little info	23	2.6
Well organized	17	1.9	Too complicated/Too much info	2	0.2
Well researched	10	1.1	Too abstract and theoretical	1	0.1
Important/relevant topic	41	4.6	Psychology not culture	1	0.1
Scenarios			Scenarios		
Multifaceted/Interesting scenarios	84	9.4	Use different scenario	26	2.9
Appropriate scenarios	7	0.8	Unrealistic Scenarios	15	1.7
Realistic scenarios	59	6.6			
Characters			Characters		
Good cultural details and variety	20	2.2	Stereotyping	6	0.7
Choices/options were good	7	0.8	Verbiage-action mismatch	36	4.0
Character actions interesting	15	1.7	Missing nuances	3	0.3
Characters/Interaction realistic	27	3.0	Focus on wrong/bad actions	40	4.4
			Characters/interaction unrealistic	7	0.8

Overall, comments about content were few in number but the favorable ones did mention "It was well organized" and "Good support material." Additionally, 4.6% of the surveys identified the topic of culture as important or relevant with comments such as "The scenarios were surprisingly relevant to real-world operations" and " I like that it offered a new way to learn some important material." However, some students found it too simplistic and felt it contained too little information. The following were representative of the negative comments in this subarea: "I think the simulation is simplistic in what I think it is trying to accomplish" and "I would like to have a little more background information and choice in scenarios."

The second subarea, scenarios, was the focus of many of the comments. Those who commented favorably wrote statements such as "The overarching scenarios were well thought out and provided a good variety of circumstances" and "Situations seemed realistic."

Several of the unfavorable comments for the Scenarios subarea specifically mentioned adding other scenarios such as Afghanistan, Iraq, the Far East, and working through an interpreter. Many of these were not negative about the simulation scenario but identified that other scenarios might make it better. However, some comments mentioned the scenarios were unrealistic. Those comments centered on two issues. The first unrealistic scenario issue was the overall mission and setting of the simulation. The second issue was more specific, critiquing actions within given scenarios. Two representative comments included "The content was good but it is unrealistic for most Air Force Officers to be in that situa[t]ion" and "...the notion that I'd be meeting with some of these guys with AK-47s everywhere and the risk of being kidnapped is a joke, that's not how it works."

Comments regarding the mission being unrealistic either specified they personally would never be tasked to do that mission or they commented that the mission only applied to a small number of Air Force officers. The purpose of the simulation was not to train students for a specific mission but to enhance their understanding of interacting with different cultures. However, a mission the student could envision him/herself in clearly would add realism to the simulation. The problem was the wide diversity of career specialties within the Air Force and the culturally diverse settings in which military members engaged in missions around the world made it improbable to address this issue with a single mission simulation. The drawback to using multiple missions and settings would be the cost in time and money to create and maintain the more robust simulation. A lower cost solution might be to take steps to adjust the students' perspectives on what they are going to be doing. One possibility would be adding an introduction to the simulation acknowledging the specific mission within the simulation might only be applicable to a small number of Air Force officers but reminding the students the purpose of the simulation was to focus on learning about interacting with cultures, not learning how to handle security at a specific airfield. An emphasis at the beginning of the simulation could help frame the students' perspectives, allowing them to more easily look past any specific mismatches between their specialty and the mission so they could focus on, and engage more fully with, the simulation.

The second issue regarding unrealistic scenarios dealt with specific actions within some scenarios such as interacting with warlords and medicine men that some students commented was not likely to happen. Similar to the previous issue with missions, this had the potential to decrease the sense of realism and engagement that are desirable elements of simulations. This created a dilemma for simulation designers to balance specific engagement activities representing cultures distinctly different from their own with the need to make it realistic for the student. Again, adding clarifying guidance at the onset of the simulation and an additional reminder during or after the simulation might help students shift their focus from identifying how it would not apply to them to focusing on learning about cultural interaction.

Characters were the third subarea of the Content theme. Characters were an integral part of the simulation experience. The simulation decisions faced by the user involved determining which character to talk to and what to say to them. There were 7.7% favorable comments and 10.2% unfavorable comments. Representative positive comments were "Enjoyed the interaction with the characters and role based simulation" and "Individual characters were well-described with good amount of background to help understand their biases, etc." Representative negative comments were "Often the type of response selected and the way it was actually carried out by the program (choice of words, tone used) was a mismatch & significant limiting factor" and "Teaching senior leaders that bribes, witchcraft and other unethical actions are relevant to good leadership undermines the core values of service."

The negative comments merit further explanation. Two main issues were identified within the comments. The first issue was a mismatch between the choice descriptions and the avatar's subsequent words and tone. The second issue involved situations like being asked to bring alcohol into the country for the warlord or having to participate in a ritual where blood was splattered on the participants.

Regarding the first point, the comments appeared valid. The dialog choices within the simulation had very short descriptions. Once a choice was selected, the avatar would then speak to the character using one or more full sentences to execute the selected option. Sometimes the words and tone of the avatar seemed much different than the short description given in the dialog choices. The choice descriptions were kept brief to allow the user to view all four choices together on the screen. Participants suggested allowing the user to preview the actual words and tone the avatar would use for each option would help students make a more informed decision in selecting options. Although the Scene Reset function essentially provided this capability, there were two problems with relying on the Scene Reset function to resolve this issue. The first problem was many users did not know the Scene Reset function existed. This was previously identified and discussed. The second issue was the simulation was intended provide students with the sense they were actually facing that problem. Ideally, the student was trying to make decisions like he/she would in real life, thus creating a sense of responsibility for the decision. Therefore, having to change a decision could have a negative impact on the student's simulation experience by eliciting feelings of failure in his/her decision making even though the cause of the incorrect decision was the poorly phrased description provided in the simulation. Knowing the avatar's words and tone before the choice was made would allow the user to make an informed decision, helping to engender a sense of responsible for the outcome of that decision. Enabling a preview function would be technologically feasible. It could also enhance learning by better matching the user's intent with the words and tone of voice used by the avatar. This could help develop ownership of the decision and help the student explore the cultural nuances of word choice and tone of voice.

The second negative area for Characters dealt with situations like being asked to bring alcohol into the country for the warlord or being asked to participate in a ritual where blood was splattered on the participants. The simulation allowed students to select options such as agreeing to transport alcohol into the country for the warlord. If the student agreed, then initially there was a benefit but later in the simulation, there were consequences. Several students commented that the simulation seemed to encourage unethical behavior. There were ways to complete the scenario without supplying alcohol. However, if the student refused to procure the alcohol and later was unsuccessful in finishing the simulation, then an impression might develop that the unethical path was the "approved solution." Since being asked for goods or services that by American standards is unethical is possible when dealing with non-Americans, there is merit to including those situations in the simulation. However, additional efforts could be made to ensure the simulation feedback identified that the unethical behavior could have been avoided without alienating the other person. This could help students who did not identify those other alternatives learn about negotiation tactics with other cultures and, ideally, prevent creating an impression that engaging in unethical and illegal behavior is the "approved solution."

Theme 5: Overall

The Overall theme accounts for the largest number of favorable comments. As shown in Table 26, there were 42.7% favorable and 8.6% unfavorable comments. This theme had two major subareas and one minor subarea. The first major subarea was the students' general perspectives on using the simulation and the second was what they perceived they learned by using the simulation. The third subarea dealt with commenting on completing the simulation.

Table 26

Original Simulation: Overall Comments

Liked Most	# of Responses	% of Surveys	Liked Least	# of Responses	% of Surveys
Total Overall	384	42.7	Total Overall	77	8.6
Perspective on the Simulation			Perspective on the Simulation		
Different/Innovative approach to teaching	62	6.9	Another method would be better	21	2.3
Good learning tool	13	1.4			
Great concept	4	0.4			
Enjoyed/satisfying	37	4.1			
Interesting/impressed	25	2.8			
Compelling/engaging/ entertaining	15	1.7			
Challenging/Game feel	17	1.9			
Learned from the Simulation			Learned from the Simulation		
Increased cultural awareness	42	4.7	Not useful to me personally	24	2.7
American ≠Universal values	16	1.8	Waste of time	32	3.6
Practice intercultural negotiation	35	3.9			
Consider multiple viewpoints	4	0.4			
Dealing with deceptiveness	3	0.3			
Provoked thought	18	2.0			
Think outside the box	18	2.0			
Other			-		
Glad to complete it	76	8.5			

The first subarea involved perspectives on the simulation that contained a lot of compliments on using a different approach to learning rather than just readings and a test. This might just reflect an appreciation for simply trying a different approach rather than complimenting the usefulness of the simulation. However, 94 surveys commented on positive feelings about accomplishing the simulation, referring to the activity as fun, enjoyable, engaging, and challenging. Representative positive comments included "I love the interaction rather than simply reading" and "It was a more engaging method of relating information and it did force me to consider different ways of tackling each problem." There were 21 unfavorable comments for this subarea that suggested using other methods such as case studies, online discussions, or a classroom environment to teach the material. A representative comment was "Case studies of actual events/experiences would have been a better way to teach such material than multiple choice role play[ing]."

The second subarea dealt with what users said they learned from the simulation. There were 156 favorable comments and 56 unfavorable comments. Representative favorable comments included "I learned a lot from the simulation and to think outside the box," "It did a good job of encouraging the player to look at situations from multiple viewpoints and to try to discern each individual's own interest," and "It was stimulating and provoked thought and pushed some African Tribal culture that I was total[l]y unaware of." The favorable comments specifically cited areas the simulation was designed to teach. This supported a case for the simulation being effective.

There were 56 negative comments for this subarea. Representative negative comments were "It did not seem relevant to me and my career path," "I just spent a year in Iraq advising the Iraqis...I should not have had to do this simulation," and "This exercise was a huge source of frustration and had no relevance...it only served to waste a significant amount of my time." The negative comments had two broad issues. The first issue was some students felt the information was not useful to them personally. Generally this was because they either did not anticipate deploying overseas or they already knew the information because they had spent many years overseas. Comments in this category did not specifically convey any problems with the simulation itself but that it was not useful to those specific students due to their career specialty or background. The second issue involved vague comments that the simulation was a waste of time.

Some respondents added they did not learn anything from the simulation. It was difficult to determine whether these comments were due to the students already having intercultural experience or the students felt the simulation was not an effective learning tool. These 36 unfavorable comments should be weighed against the 136 comments citing specific areas that were learned. Additionally, several of the unfavorable comments were combined with comments about the frustration caused by repetition and resorting to trial and error to get through the simulation. The comments probably reflected students who did not learn from the simulation but the cause might have been specific problems that could be fixed and so should not be taken as condemning the approach of using educational simulations.

The third subarea was Other, which had 76 favorable comments about completing the simulation. Even though these were in response to the favorable question, they might not all reflect a favorable perspective on the simulation. Some were clearly positive, noting the satisfaction they felt from completing the challenging simulation. However, many of the comments to the "What did you like the most about the simulation" question simply stated "Completing it," which could imply a negative aspect. Some comments were clearly negative and were also accompanied by unfavorable comments including “it was a waste of time.” Because of the mixed meanings within the “glad to complete it” category, no insights should be drawn for this subarea.

Summary of Comments
Concerning Simulation

The comments provided a more descriptive look at the students’ experiences in using the simulation than Likert-scale questions could provide. However, since students only commented on the one or few things they liked most or least about the simulation, these comments did not provide a complete picture. Analysis of the comments identified adjustments that could and probably should be made to fix specific problems identified in some subareas of the Preparation and Completion and Quality of Simulation themes but the remaining aspects of those themes worked reasonably well.

The insights from the Simulation Game Play theme identified the need to make the simulation as intuitive as possible and to have a default file name for saving progress. Additionally, the approach of a multiple branch simulation should consider how onerous the consequences are when the user is required to go back a couple of scenarios to change a decision. The idea of creating a longer forward path that provides an opportunity to correctly apply the lesson in a new situation and ultimately complete the simulation should be considered. Finally, hints and feedback seemed to be missed or ignored by some students. Presenting hints and feedback in additional formats to complement the text version might garner more attention from the student, which could be more effective.

Analysis for the Content theme indicated the simulation was well researched and realistic, providing a rich variety of scenarios and character types. However, the short descriptions for the choices did not always match the subsequent tone and words of the avatar. This necessitated changing to a different choice, which sometimes impacted the student's sense of ownership of the decision. Additionally, the problems with saving progress and resetting the scene exacerbated this issue by making recovery from a bad selection potentially very frustrating. An option to preview the words and tone of each selection could mitigate this issue. The need to preview the avatar's response could be particularly important for simulations dealing with culture and communication since nuances of tone, inflection, and word choice could make a big different on how the messages were received.

Analysis of the Overall theme indicated the use of a simulation was perceived as a good approach to learning. The simulation contained characteristics of game play such as engagement and still provided a learning experience.

The VEST simulation analysis, which is presented next, provides additional insight into each of these themes. A comparison between the two simulations is covered with each theme and a comparison of insights is included in the comment summary section.

Visual Expeditionary Skills Training Comment Analysis

Comments from the open-ended questions of most and least liked elements of the VEST simulation were analyzed using the same major structural themes as the original simulation analysis. The subareas were based on VEST specific responses. The purpose was to gain a broad understanding of the student experience with VEST and to add to the insights derived from the original simulation comment analysis. Although the discussion does not cover every subarea, all the subarea data are presented in tables to provide a more comprehensive picture of the comments. The analysis identified specific issues regarding the VEST simulation and insights into using simulations for education. Overall, there were larger percentages of favorable responses than unfavorable responses for every theme except Preparation and Completion. Response rates for each major theme are listed in Table 27. Percentages are based on 265 total surveys. Using percentages enabled a more direct comparison to comments from the original simulation.

A breakout of each major theme and comparison to the original simulation comments are provided in the following paragraphs. Fifty-two surveys had a blank or an inconsequential comment (such as "Nothing") for the question asking what the student liked the most about the simulation. There were 78 surveys that had a blank or an inconsequential comment for the question asking what the student liked least about the simulation. Additionally, two comments for the favorable question were negative and 25 comments for the unfavorable question were positive. Those comments were added to the appropriate columns when they were distinct from the comments that specific survey had for the other question. Additionally, four negative comments were from students who had attempted both simulations, completed VEST, and chose to comment on the problems with the original simulation. Those comments were consistent with previously discussed comments from the original simulation.

Table 27

Visual Expeditionary Skills Training Comments by Major Theme

Liked Most	# of Responses	% of Surveys	Liked Least	# of Responses	% of Surveys
Preparation and Completion	3	1.1	Preparation and Completion	83	31.3
Simulation Game Play	68	25.7	Simulation Game Play	47	17.7
Simulation Technical Quality	72	27.2	Simulation Technical Quality	4	1.5
Content	134	50.6	Content	34	12.8
Overall	86	32.5	Overall	29	10.9
Blank or "Nothing" comments	52	19.6	Blank or "Nothing" comments	78	29.4

Visual Expeditionary Skills Training
Theme 1: Preparation and Completion

The theme of Preparation and Completion encompassed those elements of initial setup and getting credit for completion that were not specific to playing the simulation. The question asked, "What did you like the most/least about the simulation?" Responses in this category were factors outside the simulation itself, yet those factors could impact the effective use of the simulation. A lower rate of comments could be considered better, even for the positive question, because positive comments here meant this area was more worthy of comment than the simulation itself. As shown in Table 28, there were minimal positive comments and the two comments in the No Technical Problems subarea might have been participants who tried the original simulation but experienced technical problems so the difference was noteworthy to them. The favorable comment numbers were comparable to the original simulation, which also had 1.1% positive comments in this area. On the negative side, bandwidth appeared to be a big issue. While the original simulation had 15.4% negative comments for this theme, VEST had 31.3% negative comments with a majority being bandwidth-specific issues. Representative negative comments were "Not always the case but, I did this simulation remotely while deployed and the download speeds took forever in some cases" and "Was difficult at times due to bandwidth issues at my place of work." Many of these comments mentioned being in a deployed location where bandwidth capability was limited. The web-based nature of the simulation and the high resolution video within the simulation required a broadband connection for the simulation to operate smoothly. Suggestions by the participants such as sending DVDs or creating lower resolution videos could mitigate this problem.

Although commented on in small percentages, there were issues with the completion process of ensuring credit for the simulation. Part of this stemmed from the fact the simulation was initially developed as a non-credit course without a completion certification process integrated within the design of the simulation. Consideration should be made for incorporating some type of completion certification at the end of the simulation. This could be simply the ability to print a certificate but having this option would have provided flexibility for how the simulation was implemented.

Table 28

Preparation and Completion Comments for Visual Expeditionary Skills Training

Liked "Most"	# of Responses	% of Surveys	Liked "Least"	# of Responses	% of Surveys
Total Preparation and Completion	3	1.1	Total Preparation and Completion	83	31.3
Preparation			Preparation		
No tech problems	2	0.8	Bandwidth	51	19.2
			Needed government computer	7	2.6
			Problems with initial directions	4	1.5
			Problem with small laptop	1	0.4
Completion			Completion		
Could retake the test	1	0.4	Not clear when done	5	1.9
			Instructions for completion	7	2.6
			Test	8	3.0

Visual Expeditionary Skills Training
Theme 2: Game Play

The Game Play category specifically looked at the operation of the simulation. This included the ability to navigate within the simulation, the design behind the simulation and feedback, and hints within the simulation that help keep the user progressing to completion. Compared to the original simulation comments that were

21.8% favorable and 73.4% negative, VEST had 25.7% favorable and only 17.7% negative. The breakout of VEST Game Play comments is provided in Table 29.

Table 29

Game Play Comments for Visual Expeditionary Skills Training

Liked Most	# of Responses	% of Surveys	Liked Least	# of Responses	% of Surveys
Total Game Play	68	25.7	Total Game Play	47	17.7
Navigation			Navigation		
Easy to navigate	4	1.5	Difficult to save progress	4	1.5
Quick/easy to get through	7	2.6	Difficult to navigate, awkward between modules	21	7.9
Design			Design		
Seeing how choices played out	26	9.8	Had to view entire scene rather than reading it	1	0.4
Good selection of choices	2	0.8	Can't tailor to personal experience	1	0.4
Multiple pathways	1	0.4	No true interaction	1	0.4
Interactive	12	4.5	Repetition	10	3.8
Feedback/Hints			Feedback/Hints		
Feedback was good	16	6.0	Feedback for wrong choices even when correct	7	2.6
			Wanted more feedback on why wrong was wrong	1	0.4
			Did not like some feedback	1	0.4

Representative positive comments were "Seeing the consequences of my choices played out in the video" and "The feedback explaining why choices were correct or incorrect and the impact on cultural relations if the incorrect choice was made." Representative negative comments included "Th[e] transition points are not clear...the simulation should automatically take you to the next "module" without user prompt," "Lots of repetition between the Afghanistan and Iraq scenarios," and "Having to go through the reasons why the other answers were wrong when a correct answer was selected."

On the positive side, participants commented favorably about visually seeing the impact of their choice within the situation regardless of whether their selected choice was right or not. They also liked the feedback given on what was wrong or right with the decision. While 4.1% of participants gave positive comments about navigation, 9.4% gave negative comments. Specifically, the transition between different modules within the simulation was noted as "not intuitive." Also, 3.8% of surveys identified repetition as a problem. This seemed generally to refer to the repetition between the Afghanistan module and the Iraq module, which had a few related points due to some similarities between Muslim cultures in that region of the world. Finally, although 6.0% thought the feedback in the simulation was good, 2.6% did not like getting feedback that mentioned why the wrong choices were wrong even when they selected the correct answer. The simulation only played the video result of the choice selected but textually conveyed the reason why the other choices were not as good. The low number of negative comments should not detract from using this approach.

Visual Expeditionary Skills Training
Theme 3: Simulation Quality

Comments regarding the quality of the simulation provided a stark contrast between VEST and the original simulation. The breakout of VEST Simulation Quality comments is provided in Table 30. Whereas the original simulation had 8.6% positive and 10.8% negative, VEST had 27.2% positive and only 1.5% negative. Specifically, comments about the high quality simulation or high quality video were made on 25% of the surveys. Representative comments regarding simulation quality were "I was really surprised at the quality with which this sim was made. Well done!," "It was a very professionally designed simulation," "I used VEST...production was outstanding. Good quality video and story was engaging," and "Good acting and effects--definite surprise and ‘wow’." Many surveys clearly expressed a very positive view of the simulation quality. There were no trends among the very small number of negative comments.

Table 30

Simulation Quality Comments for Visual Expeditionary Skills Training

Liked Most	# of Responses	% of Surveys	Liked Least	# of Responses	% of Surveys
Simulation Quality	72	27.2	Simulation Quality	4	1.5
Great/Good GUI/Graphics	5	1.9	Graphics boring	1	0.4
Great/Good Voice/sounds	1	0.4	Rehearsed, bad acting	2	0.8
High Quality, Professional, "Wow"	55	20.8	Encountered glitch in learning module	1	0.4
Quality Video	11	4.2			

Visual Expeditionary Skills Training
Theme 4: Content

Visual expeditionary skills training comments on Content also compared favorably to the original simulation. The original simulation had 32.6 % favorable comments and 17.8% negative. The VEST had 50.6% positive and 12.8% negative as shown in Table 31. Representative positive comments were "The material was relevant and most importantly useful," "VEST: Very lifelike; realistic actors in very credible situations, good explanations of cause/effect of decision making," and "Variety of paths and detail within the scenarios as well as the different outcome videos based on choices made." Representative negative comments were "The sim limited the culture to only a single ethnic group and region of the world. The military deals with a variety of cultures and should be exposed to the differences of many of the main ones," "I already knew the information," "Although the simulation was excellent I felt it was irrelevant to me because I'm not deploying to Iraq or Afghanistan any time soon," and "(I did VEST) I felt the simulation could have been more challenging. At this level of learning, I expected more complex scenarios and more decision points that would test and reinforce my learning."

Table 31

Content Comments for Visual Expeditionary Skills Training

Liked Most	# of Responses	% of Surveys	Liked Least	# of Responses	% of Surveys
Total Content	134	50.6	Total Content	34	12.8
Overall			Overall		
Up-to-date content	5	1.9	No new information	2	0.8%
Well organized/planned	6	2.3	Need better background material	3	1.1
Important topic/relevant	30	11.3	Not in depth enough	7	2.6
			Not relevant, unlikely to be in this situation	4	1.5
			Different than what Army teaches	1	0.4
Scenarios			Scenarios		
Multifaceted/interesting scenarios	20	7.5	Scenarios too similar	1	0.4
			Unrealistic Scenarios	1	0.4
Realistic Scenarios	52	19.6	Want other scenarios	12	4.5
				12	
Characters			Characters		
Good cultural details and variety	11	4.2	Character full of himself	1	0.4
			Canned responses	1	0.4
Choices/options were good	5	1.9	Brown nosing	1	0.4
Characters/Interactions Realistic	5	1.9			

It is interesting that the percentage of comments identifying the topic as important or relevant topic more than doubled with VEST: 11.3% versus 4.6% for the original simulation. There is no reason to think if a question had been specifically asked about the importance of culture there would be any difference between surveys from original simulation and from VEST. Since these were open ended questions, this was probably a reflection on what issues the participants chose to comment on rather than a difference in perceptions on the importance of culture. The high quality immersive nature of VEST as well as the setting of the scenarios in Iraq and Afghanistan, which were active war zones while the surveys were being sampled, might have caused students to more readily identify the importance of the topic on the surveys.

Some of the actual tasking had similarities between the two simulations. The VEST required the student to negotiate with a local sheikh to settle a dispute and the original simulation required negotiating with the local chief to settle a dispute. However, 19.4% of surveys commented on the realistic scenarios for VEST versus 6.6% of surveys for the original simulation. This might be due to the high quality and live actor video and settings within war zones previously mentioned. However, it could also be the students felt there was a richer, fuller context presented within the VEST videos that made the scenarios more realistic.

The largest negative subarea within the Content theme was the 4.6% of comments that mentioned adding a difference scenario. As previously discussed, asking for other scenarios was not necessarily a negative comment and might in fact be a positive comment. The request for alternate scenarios might be a reflection of the specific locations or jobs those students were in such as being stationed in the Far East or working in SOUTHCOM (focused on South America). It could also be that the two scenarios within VEST had a lot of cultural similarities and students wanted to see a non-Muslim, non-Middle East scenario added to VEST to enhance their exposure to difference cultures. One survey suggested making a female version of the scenario. This simulation did not make accommodations for the gender of the participant. Although only one survey mentioned the issue, serious consideration should be made to adding a second set of videos with a female leading the effort. Although a second set of videos for each scene would increase the time and expense of the simulation, it could also provide a more appropriate experience for the female students by capturing the nuances of gender-based attitudes within that culture.

Visual Expeditionary Skills Training
Theme 5: Overall

As shown in Table 32, only 32.5% of the VEST surveys had favorable comments for the Overall theme, which was fewer than the 42.7% of surveys for the original simulation. Additionally, the favorable comments focused more on perspectives regarding the simulation experience and less on describing what was learned or on the Other subarea. Positive comments included "It was addictive, I kept wanting to finish the current stage just to see what would happen," "The VEST simulation was the best computer-based training I've ever done," and "VEST was a wonderful program. I had a hard time stopping the scenarios because I enjoyed them so much. VEST is a great learning experience!" Negative comments were relatively consistent between the two simulations for the Overall theme with 8.6% providing negative comments on the original simulation and 10.9% providing negative comments on VEST. Negative comments included "The simulation is extremely time consuming," "Interesting insights, however almost entirely redundant with knowledge I already had," and "Overall, the entire experience was an almost complete waste of my time."

Table 32

Overall Comments for Visual Expeditionary Skills Training

Liked "Most"	# of Responses	% of Surveys	Liked "Least"	# of Responses	% of Surveys
Overall	86	32.5	Overall	29	10.9
Perspective on the Simulation			Perspective on the Simulation		
Different/Innovative approach to teaching	6	2.3	Tactical rather than strategic	1	0.4
Good learning tool	7	2.6	Time consuming	12	4.5
Enjoyed/Satisfying	11	4.2			
Interesting/Impressed	15	5.7			
Best I've seen, would recommend it	6	2.3			
Compelling/Engaging/ Felt Like I was there	17	6.4			
Challenging / Game feel	1	0.4			
Learned from the Simulation			Learned from the Simulation		
Increased Cultural awareness	1	0.4	Not useful to me personally (already knew)	7	2.6
American values ≠ Universal values	1	0.4	Good if deploying but I didn't need it	6	2.3
Practice intercultural communication/negotiation	1	0.4	Waste of time	3	1.1
Corrected misconceptions	1	0.4			
Provoked thought	1	0.4			
Good learning experience	14	5.3			
Other					
Glad to complete it	4	1.5			

The lower percentage of positive responses for VEST in the Learned from the Simulation subarea (7.1% versus 15.1% for the original simulation) should be considered in context with the rest of the VEST results. These were open ended questions that did not directly ask for specifics about any particular theme but generically asked about the simulation. The larger percentage of favorable comments in the Perspective on the Simulation subarea (23.8% versus 19.2% for the original simulation) and the larger percentage of favorable comments for the Simulation Technical Quality theme might indicate those aspects captured the students' focus with regard to commenting but did not imply that less was actually learned from the simulation. It is important to consider the quantitative analysis results. Although perceived effectiveness is not identical to actual effectiveness, if students perceived the education was effective, it follows those students were perceived to have learned something from the experience, which they could have commented on if specifically asked to do so. The original simulation had an average perceived effectiveness rating of 4.048 using a scale from 1 to 8 and was deemed not effective even though many students commented on what they had learned. Visual expeditionary skills training had a perceived effectiveness rating of 5.611, which was considered effective. Despite the higher perceived effectiveness rating, fewer students commented on what they learned. This is likely due to using an open-ended format for the questions. Future research should consider writing open-ended questions for each of the five themes to elicit more specific comments for those areas. However, some students provided comments in the open-ended questions about the length of the survey being too long so decreasing the number of multiple choice questions should be considered if this suggestion is implemented.

Additionally, there were far less comments in the Other subarea (1.5% versus 8.5% for the original simulation), which contained nebulous phrases such as "Glad to complete it." Given the higher perceived effectiveness ratings for VEST and the higher percentage of comments describing the VEST simulation as engaging, this added support for the interpretation that these "favorable" comments reflected a negative attitude toward the simulation.

Comment Summary Concerning
Visual Expeditionary Skills
Training

The two themes of Content and Overall received the largest number of favorable comments for both simulation. Scenarios was the most commented on subarea of the Content theme, indicating both simulations provided plausible cultural situations. Although both simulations also had large numbers of comments for the Overall theme, the distribution of comments between the three subareas was different. The original simulation had larger percentages for liking that it was a different innovative approach to teaching, more readily identified what they learned from the simulation with phrases such as provided thought, corrected misconceptions, thought outside the box, but also had much larger percentages with the "glad to complete it" type of phrasing that might have negative connotations. The VEST comments from the Overall theme centered on perspectives about the simulation experience using words like enjoying, interesting, and engaging. Many students also commented that VEST was a good learning experience. There were far less comments for VEST in the nebulous "glad to complete it" subarea. Additionally, the Simulation Quality theme was a close third for percentage of favorable comments with the subarea of "High Quality, Professional, Wow" receiving the most comments of any subarea for the VEST simulation with 20.8% favorable.

The Simulation Game Play theme for the original simulation had the largest percentage of comments for any theme with 73.5% of surveys providing negative comments for this theme. Students identified problems with navigation and how the multiple branch design was implemented with dead-end paths that caused frustration. The VEST simulation was web-based and used high resolution video so users who had limited bandwidth encountered degraded performance of the simulation program. The subarea of Bandwidth within the Preparation and Completion theme received the largest number of negative comments for the VEST simulation. There were negative comments on 19.2% of VEST surveys regarding this bandwidth problem.

The similar and unique issues identified by the survey comments from the two simulations helped provide a more complete understanding of some of the benefits and problems with using simulations for professional military education.

Additional Analysis

In addition to the planned analysis, two other areas were identified for further analysis during this study. The first was an analysis of the completion codes generated by the original simulation and the second was a one-way MANOVA utilizing age as the independent variable. The inclusion of an additional MANOVA was based on the large differences in means for age identified in the descriptive statistics section of this chapter.

Analysis of Completion Codes

When students completed the original simulation, a completion code was generated containing the student's identification number and codes for the various decision paths the student took during their successful completion of the simulation. Decision paths reflected a series of decisions made during the scenarios. Different paths were the result of different user decisions made during the simulation. Different paths resulted in different simulation experiences. These data provided an opportunity to investigate if gender impacted the decisions made while completing the simulation. The decision paths within the simulation were not designed with regard to gender. The only difference within the simulation for gender was females had a female avatar with a female voice.

Air War College added gender data to the student completion codes and identifying student numbers were removed. A total of 2,671 completion codes were provided from students who completed the simulation between October 1, 2008 and August 30, 2010. These were not the same students who completed the surveys. Although this was an entirely different group, they were all Air War College students and the demographic makeup is similar. This section covers a description of that dataset, assumption testing, and a Chi-squared analysis of the data to determine if gender impacted the path the student took to complete the simulation.

Completion Code Dataset Description

The dataset of 2,671 codes included 2,272 males and 399 females. The percentage of females was 14.1%, which was slightly higher than the 11.8% female participation rate for survey data used in the previous analysis but consistent with the target population of 13% female. Each record contained a series of codes that identified the path that student took to get through a particular part of the simulation. For example, a student with a code of "AG1.1" took a different path based on decisions than a student with a code of "AP1.1" took. Additionally, the order of the codes also identified the path the student took to successfully complete the simulation. These codes only provided the path taken to successfully complete the simulation and did not contain information on unsuccessful attempts. The dataset was analyzed using a Chi-squared test for association by comparing the first codes in each of the records to see if male and female paths were significantly different. Then the second code was compared and so on through the eight sets of codes some students had. The number of codes varied from three to eight based on a multiple branch design where some paths to completion were shorter than others.

Chi-Squared Assumption Tests

There were three chi-squared assumptions to check. The first assumption was the two variables must be nominal or dichotomous. In this case, gender was dichotomous and the path was nominal. The second assumption was the variables must have two or more groups and the third assumption was the expected cell frequencies were greater than five (Lund & Lund, 2013). The last two were addressed during the analysis because each set of codes had different numbers of groups and sizes of expected cell frequency.

Chi-squared alpha. Eight chi-squared tests were accomplished for this analysis. To ensure the risk of Type I error was kept at 0.05, a Bonferroni correction was calculated, resulting in an alpha of .006 that was used to determine if significance was indicated for each of the eight tests.

Chi-squared analysis. The chi-squared analysis was performed sequentially looking at the first set of codes, then the second, and so on through the eight sets. The purpose was to compare whether gender was associated with different paths and therefore different choices made within the simulation. To avoid situations where the expected cell size was less than five, the paths that contained five or less were grouped together into a category of Other. Table 33 shows the raw data and the adjusted data for the first iteration of the test. In this first case, seven codes were combined into the Other category.

Table 33

Comparison of Raw and Adjusted Data for First Chi-Square Test

Raw Data			Adjusted Data		
1st Decision	Male	Female	1st Decision	Male	Female
AA1.1	8	0	AG1.1	2065	383
AG1.1	2065	383	AP1.1	140	11
AP1.1	140	11	Other	67	5
AP1.2	6	0			
BA1.1	12	2			
BA1.2	15	1			
BG1.1	11	0			
BP1.1	9	1			
CG1.1	6	1			

First code chi-squared analysis. As mentioned earlier, a check of the expected count was made for each chi-square test to ensure every cell was greater than 5. The first chi-squared test had expected counts greater than 5 for all cells with the lowest being 10.8. Additionally, Table 34 shows the observed frequencies and the percent for each of the decisions. Females chose path AG1.1 at a higher rate than men and they chose AP1.1 and Other at a lower rate.

An alpha of 0.006 was used. The Pearson chi-square *p*-value was 0.003. There appeared to be a statistically significant association between gender and the first chosen path within the simulation program.

Table 34

Chi-Square Cross-Tabulation for First Set of Codes

		Decision			Total
		AG1.1	AP1.1	Other	
Male	Count	2065	140	67	2272
	% within Gender	90.9%	6.2%	2.9%	100.0%
Female	Count	383	11	5	399
	% within Gender	96.0%	2.8%	1.3%	100.0%
Total	Count	2448	151	72	2671
	% of Total	91.7%	5.7%	2.7%	100.0%

Although chi-square can identify whether or not there appears to be a relationship, it does not provide a way to measure the strength of the relationship. To determine the strength of the association, a Cramer's V was used. Cramer's V ranges from 0 to 1.0. The Cramer's V was 0.066 with a *p*-value of .003. It appears there was a weak association between gender and the first decision code.

Second code chi-squared analysis. The second set of codes was run using the same method of grouping all counts of 5 or less into the Other group. All expected values were greater than 5 with the smallest being 10. The difference in percentages between males and females appeared to be very small for every decision as shown in Table 35. The *p*-value was 0.605, indicating no significant association was found between gender and the choices for the second code.

Table 35

Cross-Tabulation of Chi-Square Data for Second Code

		Decision							Total
		BA1.1	BA1.2	BG1.1	BP1.1	CA1.1	CG1.1	Other	
Male	Count	479	501	569	346	58	240	79	2272
	% within Gender	21.1%	22.1%	25.0%	15.2%	2.6%	10.6%	3.5%	100.0%
Female	Count	76	93	106	68	9	39	8	399
	% within Gender	19.0%	23.3%	26.6%	17.0%	2.3%	9.8%	2.0%	100.0%
Total	Count	555	594	675	414	67	279	87	2671
	% of Total	20.8%	22.2%	25.3%	15.5%	2.5%	10.4%	3.3%	100.0%

Third code chi-squared analysis. The third set of codes contained eight decision paths with greater than five occurrences. Fifteen paths contained five or less and were grouped together in the other category. Table 36 shows the actual count and percentage for each decision. As with the second code, the difference in percentages appeared to be small. The Pearson chi-square test was 0.519, indicating no statistically significant association was found between gender and the third chosen path within the simulation program.

Table 36

Cross-Tabulation of Chi-Square Data for Third Code

		Decision									Total
		BA1.4	BP1.3	CA1.1	CA1.5	CA1.8	CG1.1	CG1.2	DG1.1	Other	
Male	Count	442	318	252	33	70	708	234	63	152	2272
	% within Gender	19.5%	14.0%	11.1%	1.5%	3.1%	31.2%	10.3%	2.8%	6.7%	100.0%
Female	Count	87	63	42	10	9	114	39	7	28	399
	% within Gender	21.8%	15.8%	10.5%	2.5%	2.3%	28.6%	9.8%	1.8%	7.0%	100.0%
Total	Count	529	381	294	43	79	822	273	70	180	2671
	% of Total	19.8%	14.3%	11.0%	1.6%	3.0%	30.8%	10.2%	2.6%	6.7%	100.0%

Fourth code chi-squared analysis. The fourth code was the point where some students had already completed the simulation. There were 23 males and seven females who had completed the simulation and did not have a fourth code. This reduced the sample size for the fourth code analysis to 2,641. Eight paths contained more than five occurrences. Thirteen paths contained five or less grouped into the Other category. Table 37 shows the breakout of actual count and percentages for each decision. Codes CG1.1, DG1.1, DP1.3, and DP1.4 had larger differences between percentages, whereas the other codes only had a small difference. The Pearson chi-square statistic was 0.004, indicating a statistically significant association between gender and the fourth chosen path within the simulation program.

To determine the strength of the association for the fourth decision code difference, a Cramer's V was computed. The Cramer's V was 0.093 with a *p*-value of 0.004, indicating a weak association between gender and the fourth decision code.

Table 37

Cross-Tabulation of Chi-Square Data for Fourth Code

		Decision									Total
		CA1.1	CA1.5	CA1.8	CG1.1	CG1.2	DG1.1	DP1.3	DP1.4	Other	
Male	Count	231	45	54	463	491	668	70	28	199	2249
	% within Gender	10.3%	2.0%	2.4%	20.6%	21.8%	29.7%	3.1%	1.2%	8.8%	100.0%
Female	Count	41	9	8	95	86	90	21	13	29	392
	% within Gender	10.5%	2.3%	2.0%	24.2%	21.9%	23.0%	5.4%	3.3%	7.4%	100.0%
Total	Count	272	54	62	558	577	758	91	41	228	2641
	% of Total	10.3%	2.0%	2.3%	21.1%	21.8%	28.7%	3.4%	1.6%	8.6%	100.0%

Fifth code chi-squared analysis. The fifth code showed 177 males and 50 females had completed the simulation. They were removed and the sample size was further reduced to 2,444. There were six paths with more than five occurrences and 11 paths with five or less. Again those with five or less were grouped into the Other category. With the exception of DP1.3, all the groups had larger differences in percentages between genders for the various decisions as shown in Table 38. The Pearson chi-square statistic was 0.003, indicating a statistically significant association between gender and the fifth chosen path within the simulation program. The Cramer's V was 0.085 with a *p*-value of 0.003, indicating a weak association between gender and the fifth decision code.

Table 38

Cross-Tabulation of Chi-Square data for Fifth Code

		Decision						Total
		CG1.2	DG1.1	DP1.2	DP1.3	EG1.1	Other	
Male	Count	378	709	30	99	568	311	2095
	% within Gender	18.0%	33.8%	1.4%	4.7%	27.1%	14.8%	100.0%
Female	Count	77	131	10	19	82	30	349
	% within Gender	22.1%	37.5%	2.9%	5.4%	23.5%	8.6%	100.0%
Total	Count	455	840	40	118	650	341	2444
	% of Total	18.6%	34.4%	1.6%	26.6%	650.0	14.0%	100.0%

Sixth code chi-squared analysis. The sixth code showed 518 males and 99 females had completed the simulation. They were removed, reducing the *N* to 2,054. There were seven paths with more than five occurrences and seven paths with five or less. Those with five or less were grouped into the Other category. With the exception of EA1.1, which had almost no difference, all the other groups had larger differences in percentage between the genders as displayed in Table 39. The Pearson chi-square statistic was 0.010, which was larger than the Bonferroni corrected alpha of 0.006, so there did not appear to be a statistically significant association between gender and the sixth chosen path within the simulation program.

Table 39

Cross-Tabulation of Chi-Square Data for Sixth Code

		Decision								Total
		DG1.1	DP1.3	EA1.1	EG1.1	EP1.2	FG1.1	FP1.4	Other	
Male	Count	328	25	36	615	81	444	140	85	1754
	% within Gender	18.7%	1.4%	2.1%	35.1%	4.6%	25.3%	8.0%	4.8%	100.0%
Female	Count	66	9	6	98	27	64	20	10	300
	% within Gender	22.0%	3.0%	2.0%	32.7%	9.0%	21.3%	6.7%	3.3%	100.0%
Total	Count	394	34	42	713	108	508	160	95	2054
	% of Total	19.2%	1.7%	2.0%	34.7%	5.3%	24.7%	7.8%	4.6%	100.0%

Seventh code chi-squared analysis. The seventh code showed 1,300 males and 238 females had completed the simulation. They were removed and the *N* was 1,133. There were three paths with more than five occurrences and six paths with five or less. Those with five or less were grouped into the Other category. All decisions had large differences in percentages between genders as shown in Table 40. The Pearson chi-square statistic was 0.044, indicating no statistically significant association was found between gender and the seventh chosen path within the simulation program.

Table 40

Cross-Tabulation of Chi-Square Data for Seventh Code

		Decision				Total
		EG1.1	FG1.1	FP1.4	Other	
Male	Count	274	459	149	90	972
	% within Gender	28.2%	47.2%	15.3%	9.3%	100.0%
Female	Count	62	68	22	9	161
	% within Gender	38.5%	42.2%	13.7%	5.6%	100.0%
Total	Count	336	527	171	99	1133
	% of Total	29.7%	46.5%	15.1%	8.7%	100.0%

Eighth code chi-squared analysis. The eighth code showed 1,977 males and 331 females completed the simulation, reducing the sample size to 363. There were two paths with more than five occurrences and four paths with five or less. Those with five or less were grouped into the Other category. However, when grouped together, there were four males and three females. Because the group sizes for Other were less than five, those samples were removed from the dataset, further reducing the sample size to 356. All groups had small differences in percentages based on gender as shown in Table 41. The Pearson chi-square statistic was 0.212, indicating no statistically significant association between gender and the eighth chosen path within the simulation program. Because there were only two paths for this analysis, both variables were dichotomous. Therefore, a Fisher's exact test was also used. The Fisher's exact test produced a *p*-value of 0.297, which supported the same conclusion as the Pearson chi-square test--the null hypothesis should be rejected.

Table 41

Cross-Tabulation of Chi-Square Data for Eighth Code

		Decision		Total
		FG1.1	FP1.4	
Male	Count	231	60	291
	% within Gender	79.4%	20.6%	100.0%
Female	Count	56	9	65
	% within Gender	86.2%	13.8%	100.0%
Total	Count	287	69	356
	% of Total	80.6%	19.4%	100.0%

The eight tests resulted in three statistically significant results with relatively low Cramer's V coefficients, indicating a weak association between gender and those three decision paths. This result was likely context driven and might exist because decisions about cultural interaction were conceivably affected to a greater extent by gender than other types of decisions such as adjustments to fuel mixtures for a jet engine simulation. The association between gender and decisions made for this cultural simulation indicated consideration should be given to the potential impact of gender differences when designing these types of simulations.

Age as the Independent Variable

The descriptive statistics presented earlier in this chapter suggested a potential difference in the means for dependent variables based on age as the independent variable. A post hoc analysis for both the original simulation and VEST datasets was accomplished to examine this pattern more closely. The dependent variables remained as perceived effectiveness, usability, confidence, and the combined SAR (satisfaction, attention, and relevance) variable.

Assumption Check

The dataset was previously examined using gender and video game experience as the independent variables. This examination focused on those assumptions that were impacted by using a different independent variable. The sample size was 898. A detailed explanation of this assumption check is presented in Appendix G. A summary of the assumption results is presented in Table 42. Although the group size met the adequate sample size assumption, the 30-35 age group was the smallest group with only seven samples. Similar to the original assumption check, the only areas of concern were the univariate or multivariate outliers.

Table 42

Results of Assumption Check

MANOVA Assumptions	Univariate	Multivariate
R1. DV measured as interval	✓	✓
R2. IV were categorical	✓	✓
A1. Independent Observations	✓	✓
A2. Adequate Sample Size	✓	✓
A3. No Outliers	20 identified	3 identified
A4. Normality	✓	✓
A5. Linearity	✓	✓
A6. Homogeneity of Variance	✓	✓
A7. Multicollinearity	✓	✓

Confidence had 20 univariate outliers. The combined SAR variable had one outlier, but it was a sample that was also an outlier for confidence. There was no indication of a data entry error. The breakout of outliers is presented in Table 43.

Table 43

Univariate Outliers Broken Out by Group

Age (years)	Number of Outliers
30-35	0
36-40	5
41-45	11
46-50	1
>50	3

To accomplish the check for multivariate outliers, a Mahalanobis distance was computed and compared against the chi-square distance for the degree of freedom equal to the number of dependent variables--in this case, four. The critical value was 18.47. Three cases that exceeded this value (case numbers 253, 246, and 135) had no indications of data entry errors. Samples 246 and 135 were univariate outliers but sample 253 was not. The total number of surveys with univariate and multivariate outliers was 21.

The outliers all seemed to be valid ratings. Additionally, out of a total of 898 samples, 21 outliers were only 2.3% of the surveys. However, similar to the previous MANOVA tests, one MANOVA was accomplished with the outliers present and a second MANOVA was accomplished with the outliers removed.

One-Way Multivariate Analysis of Variance with Outliers Present

A one-way MANOVA was accomplished using the original dataset ($N = 898$) with the outliers present using participant age as the independent variable. As shown in Table 44, with the exception of the small 30-35 age group, perceived effectiveness and the combined SAR tended to have higher means for the older groups whereas usability tended to have a lower mean for the older groups. Confidence ratings had less variation across the age groups and did not have a clear trend up or down.

Table 44

Descriptive Statistics for Age

Age (years)	*N*	Perceived Effectiveness $\bar{x}$(s)	Usability $\bar{x}$(s)	Confidence $\bar{x}$(s)	SAR $\bar{x}$(s)
30-35	13	4.231 (1.74)	8.692 (3.45)	33.692 (4.63)	79.38 (18.59)
36-40	380	3.713 (1.82)	8.500 (3.40)	33.329 (5.49)	77.99 (21.71)
41-45e	314	4.089 (1.74)	8.475 (3.46)	33.697 (5.57)	82.07 (20.94)
46-50	136	4.507 (1.86)	8.140 (3.35)	33.140 (6.25)	86.99 (22.74)
>50	55	4.945 (1.68)	8.182 (3.67)	33.655 (5.26)	91.75 (20.28)

The MANOVA statistical test results are presented in Table 45 and show the Wilks-Lambda level of significance was less than 0.001, indicating age had a statistically significant effect on the combined multivariate variable. However, the partial eta squared indicated the estimated effect size was small and accounted for less than 2% of the variation. Before accomplishing follow-up testing to determine which specific dependent variables were impacted by the independent variables, a second MANOVA was accomplished to determine if outliers affected the results.

Table 45

One-Way Multivariate Analysis of Variance Statistics for Age with Outliers Present

Effect	Value	*F*	Hypothesis df	Error df	Significance	Partial Eta Squared
	.095	2124.900	4.000	890.000	.000	.905
Age	.929	4.169	16.000	2719.633	.000	.018

One-Way Multivariate Analysis of Variance with Outliers Removed

The 21 outliers were removed from the database, changing the *N* to 877. A reexamination of the assumptions was accomplished. Two additional outliers, samples 75 and 452, were identified and removed, lowering *N* to 875. All other assumptions were met.

The one-way MANOVA was run using the adjusted dataset (*N* = 875) with the outliers removed. This was accomplished to create a comparison to determine the effect of the outliers. The youngest group, 30- to 35-year-olds, did not have any outliers so the means and standard deviations remained the same. All 23 outliers were low rating outliers. Although the outliers were only in Confidence and the combined SAR variables, all the variables had slightly increased means and small decreases in standard deviations for the four oldest groups.

The MANOVA statistical analysis produced slightly different values, as shown in Table 46, but there was still a statistical significance and the effective results were the same as the previous MANOVA. The *p*-value indicated a statistical significance on the combined multivariate variable but partial eta squared results indicated the magnitude of the impact was small. These results were similar to the results of the original dataset, adding confidence to the previous MANOVA results.

Table 46

One-Way Multivariate Analysis of Variance Statistics for Age with Outliers Removed

Effect	Value	*F*	Hypothesis df	Error df	Significance	Partial Eta Squared
	.077	2595.778	4.000	867.000	.000	.923
Age	.916	4.812	16.000	2649.366	.000	.022

N = 875

Stepwise Discriminant Analysis for the Original Simulation Dataset

To further investigate the effect and identify which dependent variables were most responsible for this effect, a stepwise discriminant analysis (SDA) was accomplished. The assumptions for the SDA were the same as the MANOVA so no additional assumption testing was necessary. The same approach utilized to deal with outliers for the MANOVA was followed for the SDA. The analysis was accomplished with the outliers present (*N* = 898) and with the outliers removed (*N* = 875). The *p*-value to enter for the SDA was set at 0.05 and the *p*-value to remove was set at 0.10.

The first step of the SDA with the outliers present ($N = 898$) identified the perceived effectiveness variable. The second step retained perceived effectiveness and added the usability variable. The analysis stopped after two steps because the two remaining variables did not have a low enough *p*-value to be added to the function. The Wilks-Lambda significance for the resulting function was less than .001 so it was reasonable to assume the function explained the variation. The canonical correlation for the first function that contained perceived effectiveness was 0.227. The square of the canonical correlation was .0515, indicating perceived effectiveness accounted for approximately 5.1% of the variation. The canonical correlation function from the second step that added usability was .022. The square of the canonical correlation was .000484, indicating usability accounted for less than 1% of variation.

To check if the outliers impacted the results, the SDA was accomplished with the outliers removed ($N = 875$). Just like the previous SDA, the first step identified perceived effectiveness. However, the second step added the confidence variable. The combined SAR variable was added in the third step. The analysis stopped after three steps because usability had a *p*-value of 0.073, preventing it from being added to the function. The Wilks-Lambda for the function was less than 0.001, indicating statistical significance. The canonical correlation of the first function that contained perceived effectiveness was 0.258. The square of the canonical correlation was 0.0665, indicating perceived effectiveness was estimated to account for approximately 6.7% of the variation. This was in line with the estimated 5.1% of variation for perceived effectiveness with the outliers present. The canonical correlations for the second and third functions were 0.084 and 0.043, respectively, which provided estimates of less than 1% for both confidence and the combined SAR variable.

The SDA with the outliers present and the SDA with the outliers removed both identified perceived effectiveness with similar estimates of effect sizes. The second step of the SDA with outliers present identified usability with a significance for the "F to enter" of 0.016 versus 0.048 for confidence. When the outliers were removed, the significance to enter was 0.028 for both usability and confidence. Additionally, the Wilks-Lambda was .939 for both usability and confidence. Although confidence was selected, the difference between selecting usability and confidence was very small. Considering the removal of outliers eliminated 20 very low, yet probably accurate ratings for confidence, the identification of usability in the first SDA should not be dismissed. However, as noted, the effect size was less than 1% for usability in the first SDA. The two SDAs provided solid indications that perceived effectiveness was affected by age.

Assumption Check for Visual Expeditionary Skills Training

The VEST dataset was also used to investigate if age impacted the four dependent variables. This dataset was previously examined for meeting MANOVA requirements and assumptions using gender and video game experience as the independent variables. The dataset was reexamined focusing on those assumptions impacted by using a different independent variable. The sample size was 265. A detailed explanation of this assumption check is presented in Appendix H. The summary of the assumption results is presented in Table 47. Similar to the previous assumption check, the only assumption that indicated an area of concern was the assumption of no univariate or multivariate outliers.

Table 47

Results of Assumption Check for Visual Expeditionary Skills Training

MANOVA Assumptions	Univariate	Multivariate
R1. DV measured as interval	✓	✓
R2. IV are categorical	✓	✓
A1. Independent Observations	✓	✓
A2. Adequate Sample Size	✓	✓
A3. No Outliers	49 identified*	3 identified*
A4. Normality	✓	✓
A5. Linearity	✓	✓
A6. Homogeneity of Variance	✓	✓
A7. Multicollinearity	✓	✓

* Forty samples contained all 52 univariate and multivariate outliers.

Outliers were present for each of the four variables. Overall, there were 49 univariate outlier values across the four variables. However, some samples had outliers for multiple variables. A total of 40 samples contained all of the outliers. Of note, there were nine high rating outliers for perceived effectiveness. All of the previous outliers for the various MANOVAs had low ratings outliers. There was no indication of data entry errors. Nearly all of the outlier ratings would not have been outliers for the original simulation dataset. However, the smaller VEST dataset and smaller standard deviation led to these values being outliers. A breakout of outliers by variable is presented in Table 48.

Table 48

Univariate Outliers Broken Out by Group

Age	Perceived Effectiveness	Usability	Confidence	Combined SAR	Total Outliers	Total unique sample Outliers*
30-35	0	1	0	0	1	1
36-40	25	5	3	2	35	30
41-45	2	2	4	1	9	6
46-50	2	1	0	0	3	2
>50	0	0	1	0	1	1
				Total	49	40

* Forty samples contained all 49 outlier values.

There were three multivariate outliers: case numbers 184, 98, and 75. All three samples also contained univariate outliers. Therefore, the total number of unique samples containing outliers was 40.

The outliers all seemed to be valid numbers. Forty outliers out of 265 samples was 15.1% of the surveys, which seemed large. Similar to the previous MANOVA tests, a MANOVA was accomplished with the outliers present; if a significance had been found, a second MANOVA would have been conducted with the outliers removed.

One-Way Multivariate Analysis of Variance with Outliers Present

A one-way MANOVA was run using the VEST dataset ($N = 265$) with the outliers present using age as the independent variable. As shown in Table 49, perceived effectiveness and the combined SAR tended to have higher means for older age groups whereas usability and confidence did not have an apparent trend. The mean ratings for the VEST simulation were much higher for all ages and variables than the means for the original simulation. Additionally, the standard deviations for all VEST groups were lower than the original simulation groups except for usability in the 46 to 50 age group and the combined SAR variable for the 30-35 age group.

Table 49

Descriptive Statistics for Age with Visual Expeditionary Skills Training Dataset

Age	*N*	Perceived Effectiveness $\bar{x}$(s)	Usability $\bar{x}$(s)	Confidence $\bar{x}$(s)	Combined SAR $\bar{x}$(s)
30-35	7	5.00 (1.00)	10.14 (2.19)	39.57 (3.69)	94.57 (22.16)
36-40	106	5.41 (1.63)	11.10 (2.71)	39.45 (3.42)	98.20 (19.62)
41-45	91	5.67 (1.56)	11.18 (2.939	39.48 (3.89)	102.84 (19.04)
46-50	42	5.81 (1.74)	10.93 (3.56)	40.07 (3.43)	108.02 (17.78)
>50	19	6.26 (1.76)	11.53 (2.84)	38.89 (4.77)	108.89 (20.70)

N = 265

The MANOVA statistical test was accomplished and the results are shown in Table 50. The Wilks-Lambda significance level was 0.126 so there was insufficient evidence to conclude that age impacted any of the variables. Based on these results, a follow-up MANOVA with the outliers removed was not necessary and, therefore, was not accomplished. No additional analysis was conducted.

Table 50

One-Way Multivariate Analysis of Variance Statistics for Age with Outliers Present for Visual Expeditionary Skills Training

Effect	Value	*F*	Hypothesis df	Error df	Significance	Partial Eta Squared
	.020	3117.656	4.000	257	.000	.980
Age	.917	1.418	16.000	785.786	.126	.022

N = 265

The relatively small group size and large percentage of outliers would have resulted in less confidence in any statistically significant result from the VEST database. However, for those same reasons, the lack of a significant result did not necessarily detract from the confidence in the results from the original simulation database.

CHAPTER V

DISCUSSION

Introduction

This chapter discusses the major themes and provides an interpretation of the implications from this study. A discussion on the implications of perceived effectiveness is followed by an examination of the confidence rating issues encountered throughout this study. Next the implications of video game experience, gender, and the interaction between gender and video game experience are covered. Then the implications of the qualitative assessment are discussed, starting with an examination of the five themes generated by the study, and followed by a discussion of the findings regarding technical issues, user friendliness, and enhancements to learning. Then the implications of age of the participants as an independent variable are discussed. Finally, issues relating to surveying military personnel and recommendations for future research are presented.

Perceived Effectiveness

The first research question asked if students perceived the simulation as being effective. This question was intended to set the framework for understanding subsequent questions and their associated results. The results indicated the original simulation was not perceived as effective but the visual expeditionary skills training (VEST) simulation was perceived as effective. The first implication was derived from the fact that VEST was rated well above the midpoint. This finding implied simulations could be perceived as effective by this population. There did not seem to be a bias against the use of simulations. In fact, the comments presented many favorable views of utilizing simulations for AWC/DL education and very few views were opposed to using simulations for this level of education. This was consistent with Douglas et al.'s (2007) study, which also found simulations were viewed favorably for use in higher education. The negative views on simulations were mostly tied to specific difficulties with that simulation rather than negative views on the use of simulations in general. The negative views on the original simulation were generally tied to the issue of being "stuck in a loop" due to the multiple branch design and limited feedback. The negative views on the VEST simulation were generally tied to the simulation running poorly in locations with limited bandwidth capability. The negative views on specific aspects of the simulations are discussed later. A small number of negative comments were made about the general use of simulations. The OS simulation had 21 negative comments out of the 898 total surveys tied to the general use of simulations. Students making those comments generally indicated a preference for other methods such as case studies, online discussions, and a classroom setting rather than using simulations for learning. The VEST simulation had 13 negative comments out of 265 surveys. The 13 comments indicated simulations were too time-consuming. These comments might reflect the specific simulation or they might reflect simulations in general. The limited number of negative comments about simulations for education also helped to support the idea that simulations could be perceived as effective learning tools by this population.

The second implication of the ratings was there was a marked difference between the two simulations in terms of student experience (which is explained in more detail later in this chapter). Although the original simulation had much lower ratings for perceived effectiveness and there appeared to be several problems with the operation of the simulation, it must be kept in mind that the original simulation was the first attempt by Air War College to implement an in-depth role-playing simulation. Both simulations were developed based on technology available at the time. However, the VEST simulation--developed five years later than the original simulation--incorporated improvements in technology and programming that not only included improvements in creating the simulation but expected improvements in the capabilities of the user's computer and Internet connection to handle a more robust program. Despite the original simulation's "80s era video game" appearance some students identified in their comments, there was still positive feedback on the original simulation. In fact, based on the comments, the main detraction of the original simulation seemed to be the frustration regarding the implementation of the multiple branch design rather than poor graphics.

The final point on perceived effectiveness was the higher ratings for VEST did not necessarily mean VEST was more effective than the original simulation in accomplishing AWC/DL's learning objectives. The rating of perceived effectiveness was not a perfect measurement of actual effectiveness. Testing actual effectiveness would require monitoring the accomplishment of the simulation and then testing whether participants could actually apply their new understanding of intercultural relationships in real-world situations. While the simulation was designed to test students' ability to apply the lesson within the context of a realistic scenario, no comparable performance ratings were given. The same passing score was given to everyone who completed the simulation in a satisfactory manner. Furthermore, there was no measurement of how many attempts the student made to complete the simulation. A separate evaluation of effectiveness was outside the scope of this study. Nevertheless, military officers do routinely evaluate processes and people so it is reasonable to assume there is some association between perceived effectiveness and actual effectiveness. The ratings could provide insight into whether or not the participants "felt" they learned something from the simulation. However, the results should not be used to discount the value of the original simulation compared to the VEST simulation. The VEST simulation did not exist when this study was initiated so perceived effectiveness was never intended to be a method for comparing actual simulation effectiveness. Comments certainly indicated VEST was a more engaging experience than the original simulation, which should enhance the learning experience. However, while the VEST simulation might have been more engaging, it is possible it created a more "tactical" level experience focusing on specific actions for a specific culture rather than the more "strategic" level learning desired by Air War College that would focus on intercultural relationships in general. Rather than attempting to compare the effectiveness of the original simulation with VEST, this research focused on examining each simulation separately and determining if common issues and implications were present.

Confidence

Issues with the confidence variable were identified during many of the analyses conducted for this research. The first analysis to identify issues with confidence was the factor analysis that examined Keller's (2010) attention, relevance, confidence, satisfaction (ARCS) categories within the instructional materials motivation survey (IMMS) portion of the survey. The analysis indicated the responses to Keller's ARCS questions seemed to fall into two distinct factors. The attention, relevance, and satisfaction ratings were more similar in variances; whereas confidence seemed distinct from those three. This did not imply the ARCS model was incorrect (especially given past research). However, for this specific study, confidence was the only factor that seemed distinctly different. Because ARCS was attempting to measure motivation, it is reasonable that higher motivation generally would result in somewhat higher ratings for each of the factors so there might not be a great disparity between the four variables. The distinct nature of the confidence ratings within this study was also confirmed by nearly all the statistical measures throughout this study. Cronbach's alpha for confidence was the only one of Keller's four variables that had a large difference between the study's measurement and the established Cronbach's alpha from previous research utilizing Keller's IMMS. Also, the assumption testing for every MANOVA in the study identified outliers in the confidence ratings. This included both the original simulation dataset and the VEST dataset. While the outliers did not appear to alter the results based on running a second MANOVA to compare the impact of the outliers, their existence pointed to a possible uniqueness for the confidence variable. Some of this might be explained by the specific way in which AWC/DL operates. Every student completes the coursework on his/her own. The typical course includes readings and a test--a format similar to previous PME programs. Although more robust role-playing simulations are currently being incorporated into the curriculum for other PME programs, those types of simulations were not part of the experience for the sampled population when they completed the previous PME programs. Being less familiar with role-playing simulations coupled with the rules for the PME learning environment that did not allow students to get help from their peers might have led to more frustration and loss of confidence than might normally be encountered with the types of simulation problems identified in the comments. Although students were not allowed to work with others to complete the coursework, they could contact AWC instructors to discuss problems. However, time zone differences and limited instructor availability might have deterred contact with instructors. Using the simulation in a group setting or even in a traditional classroom environment might have led to different results for confidence. Both simulations had comments by some students reflecting very positive experiences but for other students such as those in the outlier samples, the experience was probably very poor. In part, the issue with confidence for the original simulation seemed to be the result of endless loops some students found so frustrating during the simulation. As expressed in many of the comments, they did not know how to get out of the loop and finish the simulation. This clearly could have impacted their confidence in completing that simulation. The relative isolation in which each AWC/DL student completed the program probably necessitated a more rigorous testing of the student experience than a normal environment would require since even a small problem, which could normally be easily resolved by a comment from a peer or instructor, could result in a very poor experience for the AWC/DL student who encounters it.

The VEST confidence issues seemed to have stemmed from the uncertainty of getting through the simulation due to bandwidth issues. Peer discussions would not have impacted bandwidth. Bandwidth and technical issues in general are covered in the qualitative assessment section of this chapter.

Video Game Experience

The second research question dealt with whether or not the six dependent variables of perceived effectiveness, usability, attention, relevance, confidence, and satisfaction were affected by the independent variable of video game experience (VGE). The MANOVA results indicated usability was impacted by VGE. Although the dataset had outliers, a second MANOVA with the outliers removed had a similar result, providing a strong indication of a link between the VGE and usability. Participants with high VGE seemed to find the simulation more usable. However, the MANOVA results for VGE with outliers present also indicated the confidence variable was impacted. When the outliers were removed from the dataset, the results only identified a link with usability and not with confidence. Although the outliers appeared to be valid data points, there was less support for the confidence variable being impacted by VGE. The estimate of the effect size was only 2.66%; so while there was statistical significance, the effect appeared to be relatively small. The MANOVA for the VEST simulation did not find a statistically significant impact from VGE on any of the dependent variables. However, this dataset was much smaller and consequently, the tests had less power than with the dataset on the original simulation. The link between VGE and usability was expected and was consistent with Virvou and Katsionis's (2006) study that found novices spent more time learning how to use the simulation and significantly more time incorrectly navigating within the simulation. An AWC/DL student's experience in video games appeared to make him/her more familiar with the actions and processes that were part of the culture simulations. However, the AWC/DL student population had a familiarity with computers, which every military officer uses to some degree on a routine basis. The student population also accomplished annual online training programs. The effect of VGE might have been minimized due to familiarity with computers and annual online training programs. This could mean a population without those skills and experiences would see a larger impact due to VGE.

Gender

The third research question which looked at the impact of gender on the dependent variables had similar results to VGE. Gender impact on the usability rating was statistically significant both with the outliers present and with the outliers removed. Males tended to find the simulation more usable than did females. When the outliers were present, the combined SAR variable was also statistically significant. When the outliers were removed, the combined SAR variable was not statistically significant. Similar to VGE, there was high confidence that usability was impacted by gender but the outliers made the impact of the combined SAR variable less certain. As with VGE, the estimated effect size for gender was small--around 1%. The mixed results of previous studies based on gender and the self-selection for a technology-related career for female Air Force officers led to an expectation that there would no difference. Additionally, if familiarity with computers and online training programs minimized the differences for VGE, then they would likely also minimize them for gender differences.

The MANOVA results from the VEST dataset were not statistically significant. This should not detract from the finding for the original simulation dataset because the VEST dataset was much smaller, and due to the small percentage of women in the population, the female group sizes were very small. There were only 35 females total for the VEST dataset and only 14 females who reported being experienced video gamers.

The original simulation dataset had 106 females, 51 of whom were experienced video gamers.

The issue of gender was further investigated with a chi-squared analysis for completion codes that compared actual choices made in the simulation between males and females. A Bonferroni correction was applied due to the eight iterations of the chi-squared test, so an alpha of .006 was used. The results indicated that for three of the eight decision sets, there was an association between gender and choices made. The difference in decisions should not be considered surprising given the subject of cultural awareness related closely to interpersonal communication. When dealing with some cultures, females would reasonably expect their actions and word choices to have a different impact than the actions and words of males. The difference in decision selection was likely subject-dependent and while important to consider, might not be as applicable to simulations that do not involve interpersonal communication.

The differences that seemed to exist for usability based on gender might have been due to the simulation design. The original simulation allowed the user to choose whether he/she was female or male. If female was chosen, then the voice and avatar would be female. However, all other aspects of the simulation remained the same. No additional consideration was given within the simulation for gender. The VEST simulation did not provide any options for gender selection. Regardless of the user's gender, the VEST simulation involved a lead male character facing a series of decisions as he worked to accomplish the mission. The small difference in usability ratings of the simulation based on gender might have been due to the limited, or lack of, accommodation for gender in the simulations.

Research indicated the existence of differences in usability and in choices made during the simulation due to gender. The effect size for usability was small, but the size might have been lower than it would be for the general population of this age group due to extensive computer use and computer training accomplished by Air Force officers. While the reasons for these differences were beyond the scope of this research, research findings supported a recommendation to consider gender in simulation design. Consideration should be given to the user's avatar and how it interacts with other characters. Additionally, gender differences should be considered for the range of response choices offered within the simulation. While this is likely to be dependent on the subject area of the simulation, this study supported incorporating gender consideration into the design of simulations dealing with cultural awareness and interpersonal communications.

Video Game Experience and Gender

The fourth research question asked whether there was interaction between VGE and gender regarding the six variables. While there were indications that both VGE and gender impacted usability, there was no indication to support an interaction effect between the two independent variables and usability or any of the other dependent variables.

However, the issue of VGE and gender does merit some discussion on the categorization of VGE, which was different based on gender. The median split between a rating of 4 and 5 for males determined whether male samples were categorized as inexperienced or experienced video gamers. The median split for females was between 3 and 4. It was not expected that the difference impacted any of the results because the groups were still based on more experienced versus less experienced gamers. Although consistent with Greenberg et al.'s (2010) findings involving younger populations that found females reported spending less time playing video games than their male counterparts, the results did not answer why the VGE ratings for females were generally lower than males. This question was outside the scope of this study but there did seem to be a difference in self-reported ratings for VGE based on gender. Questions 8 and 9 on the survey were used to assess the VGE category. Previous research indicated females might readily identify themselves as "playing *The Sims*" but did not consider themselves gamers (Wirman, 2014). However, the term *gamer* was not used and both questions seemed gender neutral when asking about the student's attitude toward video games and whether they played video games. The difference in ratings might be due to females actually spending less time playing less video games or it might be due to a cultural reluctance by females to identify themselves with games. Additional research would be needed to answer these questions. However, future studies should consider a potential difference between self-reporting VGE ratings by gender when developing their methodologies.

Qualitative Assessment

A qualitative assessment was used to answer the final research question that asked what the participants liked most and least about the simulation. Coupling quantitative and qualitative analysis together provided more insight into the results by revealing some of the participants' thoughts behind the ratings and helping to identify specific areas that were problematic or beneficial. The 1,871 meaningful comments were examined. From those comments, five themes emerged, which were then broken out into subareas. The results were presented in the previous chapter but a summary is provided in Table 51 depicting the five themes and percent of responses for the most liked and least liked questions. These themes helped present ideas from the remarks in a structured way. The largest percentage of positive comments for both simulations was for the themes of Content and Overall. The largest percentage of negative comments for the original simulation was simulation game play. This reflected a frustration with the simulation by students who felt stuck in an endless loop. The largest percentage of negative comments for the VEST simulation occurred in the theme of Preparation and Completion, which was primarily due to bandwidth issues. The following paragraphs provide a further discussion of these points.

The specifics for each subarea were covered in the previous chapter. However, many of the positive comments conveyed the general message that the use of simulations was a useful approach and some even considered it a desirable way to learn. Although the comment questions were open-ended rather than asking students to comment specifically on the use of simulations, 34.3% of OS surveys and 30.9% of VEST surveys had positive perspectives on the use of simulations or mentioned positive learning experiences. Many students noted they really liked the use of simulations rather than reading the large amount of written material typically used for AWC/DL lessons. Additionally, comments requesting additional scenarios for the simulations could also be regarded as positive comments about the use of simulations. Only 21 out of the 1,163 surveys indicated alternatives such as case studies would be better. Most negative comments were specific to the simulation the participant used rather than simulations in general. The 35 "waste of time" comments were not specific enough to identify whether they referred to the specific simulation used, the use of simulations in general, or simply the subject being studied. Future studies should consider using separate open-ended questions focused on each of the five themes to elicit more specific information.

Table 51

Comment Response Rate by Theme

Theme	Type of Response	% of Original Simulation Surveys	% of VEST Surveys
Preparation and Completion	Liked Most	1.1	1.1
	Liked Least	15.4	31.3
Simulation Game Play	Liked Most	21.6	25.7
	Liked Least	73.5	17.7
Simulation Technical Quality	Liked Most	8.6	27.2
	Liked Least	10.7	1.5
Content	Liked Most	32.6	50.6
	Liked Least	17.8	12.8
Overall	Liked Most	42.8	32.5
	Liked Least	8.6	10.9
Blank or "Nothing" comments	Liked Most	20.6	19.6
	Liked Least	15.1	29.4

Overall, the use of simulations seems to be worth pursuing but three general areas merit further discussion: technical issues, making the simulation "user friendly," and methods to enhance learning.

Technical Issues

Comments from both simulations reinforced the need to consider the setting and technical capabilities of the user. The geographically diverse population of AWC meant there was far less control over the technical capabilities available to students than a typical university setting. Deployed students had no control over the bandwidth at their deployed locations. They could not request more bandwidth from their Internet provider or come into a computer lab to accomplish the simulation. Technology seems to be constantly changing. While the technical issues for these simulations were specific to technology in existence at the time, the consideration given to technology will likely continue to be a critical factor in the design of future simulations.

The original simulation was designed as a stand-alone program rather than Web-based in part because of the uncertainty of bandwidth capabilities at that time for the geographically diverse student locations. The bandwidth issues identified on 19.2% of VEST simulation surveys indicated there was merit to that decision. However, the approach of using a stand-alone program had issues as well with 8.3% of comments noting issues with downloading, installing, and setting up the program. The ongoing changes made to protect military computer systems in addition to periodic software and hardware upgrades have and will probably continue to create problems for the stand-alone program approach. Current trends in transitioning to cloud-based computing and rapidly improving broadband connections should help mitigate the bandwidth issue, making it the more attractive alternative for PME. However, for the near term, consideration must be given to the limited bandwidth deployed military personnel might have available, so providing an alternative way to complete the simulation should be considered. The original simulation utilized a loadable program that could be sent on CD for the few students who could not download the program. The VEST simulation utilized high quality video within the simulation requiring broadband connection to work properly. Many students did not have sufficient Internet connectivity to run the VEST program properly. This resulted in a larger number of students needing a DVD to run VEST than the number of students who needed the CD to load the OS simulation. However, during the time period in which the survey was conducted, a limited number of DVDs were available and there were additional logistical constraints that resulted in many students having to utilize a very slow internet connection to run VEST, resulting in lengthy loading times for scenes within the simulations and sometimes an unresponsive simulation, degrading the simulation experience. For simulations that are Web-based and also require a large bandwidth capability to run properly, DVDs should be available as an option. However, some modern computers no longer have CD/DVD drives so utilizing a DVD as the only alternative to streaming the high resolution video simulation might leave some students unable to complete the simulation. To mitigate this issue, a lower resolution version could be provided on the website for users encountering problems with the higher resolution version. The high quality video experience of the VEST simulation seemed beneficial and should not be discarded on the basis of bandwidth. However, easily accessible alternatives are needed to ensure all students can access to the simulation.

User-Friendly

The comments identified a general need to make the simulation more user-friendly. The term *user-friendly* for this discussion is defined as having easy to follow but hard to avoid instructions, having intuitive or (when appropriate) automated user controls, and having clear feedback.

There were 55 specific negative comments about initially learning how to use the OS simulation. These specific comments accounted for 6.1% of the OS surveys. However, comments from other themes, such as 57 students (6.3% of OS surveys) commenting that save points caused repetition because the save points forced the student to go back several scenes to the beginning of the scenario rather than to the beginning of the current scene, also indicated issues with the initial directions. Those 57 students either did not know that a scene reset feature was available or did not know how to use it. The reset feature was covered in the initial directions. The initial instructions for VEST had only four negative comments or 1.5% of VEST surveys, indicating it was not a common problem. The instructions for the OS simulation did contain a lot of information that would have been helpful to the students, but it appeared from the comments some students either did not read or just skimmed the associated 45-page instruction manual. Additionally, based on problems described in the comments, it was likely many students did not use the OS simulation tutorial. The tutorial was not integrated into the flow of the simulation. From the main menu, students could immediately start the simulation without accomplishing the tutorial. The tutorial was completely separate from the simulation scenarios and only accessible from the main menu page. The tutorial stepped students through the simulation controls and navigation actions but did not accomplish any culture-related tasks. While efforts were clearly made to ensure the necessary instructions were available, consideration should be given to structuring the simulation to make sure the instructions are not just available, but unavoidable. Embedding the tutorial within the first scenario would ensure all students were exposed to the information. This is similar to the approach now utilized by many video games. This approach would also allow seldom used navigation functions, such as the discussion with a second character used in the original simulation, to be embedded within the applicable scenario. The embedded tutorial might be even more critical for learning environments such as AWC/DL where there is no peer interaction and limited instructor accessibility.

There were 66 negative comments (7.3% of surveys) for user navigation with the OS simulation although they did not elaborate on specific navigational issues. An additional 9.2% of negative comments dealt with saving and resetting the simulation. There were 26 negative comments dealing with saving progress and the 66 negative comments mentioned in the previous paragraph dealt with resetting the scene. The negative "resetting the scene" comments were noted above; but in addition to better initial instructions, the problems with resetting the scene could also be resolved with a more intuitive design for the user interface. Although having smaller numbers, there were also issues with knowing when the simulation was completed and what to do with the simulation once it was completed. There were 20 negative OS comments (2.2% of surveys) and seven negative VEST comments (2.6% of surveys) regarding students not knowing what steps to take when they finished the simulation to gain credit from AWC for completing the simulation.

The original simulation did have some intuitive navigational functions such as arrow keys to move left, right, forward, or backward. However, as the comment numbers above indicated, the least intuitive functions seemed to be (a) saving progress, (b) resetting the scene, and (c) generating the completion code. The save function was not directly accessible while playing the simulation. It required the user to back out of the current scene to use the save function. Additionally, the program required the user to input a file name into an area of the screen not readily identifiable as a text field. There was no default file name and selecting the save function without inputting a file name did not generate a popup error message reminding the student to input a file name. The function of resetting the scene to allow alternative choices to be made was also not readily accessible while playing the game. Both of these functions should be accessible via clearly marked buttons from any of the normal screens. Additionally, there should be a default name for saving the progress and a clearly identifiable box to input a different name for the file. Both of these issues led to frustration for many students, detracting from the simulation experience, and, in the more extreme cases, seeming to negate learning entirely as users resorted to trial and error to merely get through the simulation rather than analytically thinking through the issues and identifying a logical approach. The interface for verifying the user completed the simulation was problematic. The original simulation required the user to click on a "Send Data" button to create a code the user could then paste to an AWC/DL website. Instructions for what to do with the code were not part of the 45-page instruction guide--they were posted on the AWC/DL website. Several comments addressed the problem of dealing with the completion code. Either incorporating an automated email with the code or embedding the instructions within the simulation could have alleviated many of these problems. The VEST simulation was not originally designed for AWC/DL use. As a result, it did not have the embedded proof of completion AWC needed. A separate assessment was used instead. Both simulations could have benefited from a better and potentially automated system for verifying completion that would minimize the level of effort required by the user and instructor to complete this essential task.

Survey comments from both simulations identified the need for clear feedback especially for identifying and guiding students when significant points had been achieved within the simulation. Feedback during and at the end of the simulation was also a source of many negative comments. There were 113 negative comments (12.6% of surveys) for the OS simulation regarding feedback. Those comments identified the need for better feedback, better hints, and also noted issues with reaching a point in the OS simulation where the student could not take any additional action but where there was also no indication it was necessary to restart the scene or scenario to try a different approach. Additionally, 33 OS simulation comments (3.7% of surveys) identified a problem with students not knowing when they had successfully completed the simulation. Combining both issues, the OS simulation had 146 negative comments (16.3% of surveys) for feedback but only 19 positive comments for feedback (2.1% of OS surveys). Compared to OS, the VEST simulation did not seem to have a serious problem with feedback. There were nine negative comments regarding feedback for the VEST simulation (3.4% of surveys) and five comments about not knowing when the simulation was complete (1.9% of surveys). Combining both issues, the VEST simulation had 14 negative comments (5.3% of surveys) but there were 16 positive comments for feedback (6% of VEST surveys).

The two simulations differed on when and how they provided feedback. The VEST was designed to provide immediate specific feedback and automatically provided an opportunity to re-accomplish the scene. The OS simulation did not provide feedback until the end of the scenario. Additionally, the OS simulation provided generalized feedback rather than specific feedback and did not provide an immediate opportunity (button to press) to reset the scene or scenario. The intent was to get the student to reflect on what actions they took that might have caused the problem. It was not within the scope of this research to identify which approach (specific or general feedback) provided the better learning opportunity, but this research did indicate the OS experience for the student might have been improved by providing more feedback or hints during the scenario. Additionally, the OS simulation experience would likely have been better with a "try again" button appearing after feedback was provided regarding unsatisfactory performance in the scenario. Both simulations could have improved how they notified students that sufficient progress had been accomplished to receive credit for accomplishing the simulation. Comments for the original simulation noted some students did not realize when they had successfully completed the requirements of the simulation and continued to spend time playing optional scenarios thinking the scenarios were required. The simulation did not provide clear guidance a satisfactory level of achievement had been reached. While accomplishing the optional scenarios might have added to learning, students felt misled into accomplishing those scenarios, thus creating frustration. Several students who completed the VEST simulation noted guidance on transitions between modules was not clear. This was not due to a poor design approach; rather, the VEST simulation was not originally designed for use in the AWC/DL program. The original intent was for users to accomplish the one module appropriate to their upcoming deployment. Comments regarding both simulations illustrated the need to ensure users remained well informed of progress via clear feedback and guidance as they proceeded through the simulation.

While none of the three issues discussed in this user-friendly section dealt directly with learning the material, they all created distractions from the learning process. By improving the design, the simulation could minimize the time users spend dealing with these problems, allowing students to focus more of their time and effort on learning the material. This might be especially important in an academically isolated environment like AWC/DL.

Methods to Enhance Learning

The choice of the underlying design for both simulations was made deliberately to enhance learning. The original simulation used multiple branches. One of the reasons was to have a structure that forced students to think through the situation to resolve the problem rather than just guessing randomly, getting feedback, changing the choice, and going on to the next scenario. However, as some of the feedback indicated, a number of students still used trial and error rather than thinking through the problem to finish the simulation. This seemed to be generally due to frustration with the original simulation and the perception of being stuck in "endless loops" with no way to move forward. The "railroad" type design of the VEST simulation made it easy to simply guess your way through the simulation, yet many students reported feeling like "they were there" making the decisions. So despite using a design that effectively enabled the user to successfully resort to "trial and error," the VEST simulation created a sense of ownership of the decision and a desire to analyze the choices to make the right decision the first time. Using the videos and very realistic scenarios plus the small amount of time it took to recover from a bad decision seemed to keep the participants engaged with the simulation and minimized any time benefit to using a trial and error method. This was consistent with Norman's (1993) description of high motivation due to an engaged state of focused attention. The level of engagement generated by the VEST experience seemed to uphold at least some of the desired thinking envisioned with the original simulation even without the forcing mechanism of non-specific feedback and multiple decisions to sort through. Because only perceived effectiveness was measured, the study could not conclude what level of learning was actually accomplished for each of the simulations. However, the VEST simulation, with its high quality, live-actor videos, engaged the learner and seemed to create a very positive learning environment. This did not indicate the railroad simulation was better than the multiple branch simulation; rather, truly engaging users in the simulation might be a better approach than creating structural obstacles for preventing "trial and error" accomplishment of the simulation. Problems that led to the frustration with the multiple branch design of the original simulation could be mitigated by providing recovery paths instead of dead ends. Additionally, improvements to the save and reset functions mentioned previously could also enhance the user's simulation experience, potentially increasing his/her engagement.

Impact of Age

Post hoc analysis examining the age of the participant as an independent variable was performed on the research data. Although this was not part of the originally planned research, the clear trend within the descriptive data indicated there might have been a significant difference in ratings based on age. There were 21 samples containing outlier values in the original simulation dataset. Analysis with the outliers present and with the outliers removed found a statistical significance for the perceived effectiveness variable. Older students tended to provide higher ratings for perceived effectiveness. Additionally, the usability variable was identified as statistically significant with the outliers present but not with the outliers removed. The "F to enter" value for usability was low enough to be included during the second step but with the outliers removed, the SDA favored the confidence variable because it had a lower "F to enter" value than usability. The fact the "F to enter" was low enough to qualify during the second MANOVA added to the confidence that usability was probably affected by age as well. However, unlike perceived effectiveness, usability tended to have lower means for older age groups.

The VEST dataset was much smaller than the original simulation database but had almost twice as many outlier surveys. Outliers were present in 15.1% of the VEST surveys but in only 2.3% of the original simulation surveys. Additionally, the smallest group size was seven. Although that met the assumption for MANOVA analysis, it might have contributed to the lack of a statistical significance from the MANOVA with the VEST dataset. Given the large numbers of outliers and small group size, the lack of a significant result should not detract from confidence in the results of the original simulation analysis.

It might be that older students were more impressed with the technology of the simulation than younger students. This might reflect perception differences between those individuals who grew up before computers were generally available for educational purposes and those who grew up using computers for education. The former might be more impressed with technology in the educational setting, which might have resulted in higher ratings. The lower ratings for usability from the older age groups might also be due to less familiarity with technology in the learning environment. Additional research would be needed to confirm if rating differences existed; but if they do, those differences could potentially impact any research that utilizes self-reported data to investigate technology-related questions.

Web-Based Surveys for Military Members

The response rate for the survey was satisfactory but several factors should be considered when surveying military personnel. Some skepticism about the survey invitation was evidenced by frequent emails to the researcher confirming the authenticity of the survey. This might have been partially due to the annual information protection and information assurance training required for Air Force members that is intended to instill caution when receiving unsolicited emails with links. It is likely some invitees, unsure of the survey's authenticity, simply deleted the survey email rather than follow-up with an email or phone call to the researcher to verify the survey's status. Clearly, the inclusion of a military email address to make it easier for invitees to check the authenticity is important. Additionally, frequent change of assignment locations and frequent deployments to austere locations could potentially be a hindrance to getting an invitation to the student. Although the military has shifted to permanent email addresses, email addresses maintained by AWC are based on information provided by the students and thus included many civilian email addresses. Some, but not all, civilian email addresses could become outdated after a move, especially if the email account was through a local Internet service provider. Fortunately, the requirement to survey only newly completed students ensured most of the email addresses were still viable. If there is a longer period of time between graduation and the survey, then the response rate could be lower due to outdated email addresses. It is expected that a paper-based survey delivered through the mail system would have generated less skepticism regarding authenticity but might still result in a lower response rate due to less surveys reaching students because of the frequent moves by military personnel.

Future Research

Results of this research pointed to the need for additional exploration of issues regarding confidence, the types of simulations in which choices need to consider the user's gender, and the impact of age on perceived effectiveness. This study noted confidence seemed to be an issue for this target group, which was required to accomplish their studies in isolation. Further investigation of this topic could identify ways to better mitigate the effect. The chi-squared analysis of the completion codes indicated differences in the choices made by females compared to males for the OS simulation. Additional research might help identify other study areas in which designers should consider gender when identifying choices for the user to make within the simulation. Finally, the post hoc analysis indicated older users perceived the simulation to be more effective than younger users, but there was also some indication that older users gave lower ratings for usability. Additional research should be conducted to investigate these issues.

Conclusion

This study found statistically significant impacts due to video game experience and gender on usability but the estimates of the effect sizes were small. There was insufficient evidence to support any impact on the other dependent variables: perceived effectiveness, attention, relevance, confidence, and satisfaction. However, the study did find confidence ratings seemed to have a uniqueness within this relatively isolated learning environment. The learning environment prevented peer collaboration and added time-zone induced difficulties in contacting instructors, perhaps making it more critical to identify and resolve technical and structural issues with the simulations to minimize the negative impact on the student's simulation experience.

REFERENCES

2011 USAF Almanac. (2011). *Air Force Magazine, 94*(5), 38-56.

Air University. (2006). *Air University catalog academic year 2006-2007*. Maxwell Air Force Base, AL: Air University Press.

Alessi, S. M., & Trollip, S. R. (2001). *Multimedia for learning: Methods and development* (3rd ed.). Boston: Pearson.

Annetta, L. A. (2008). Video games in education: Why they should be used and how they are being used, *Theory Into Practice,47*(3), 229-239.

Barab, S., Zuiker, S., Warren, S., Kickey, D., Ingram-Goble, A., Kwon, E., …Herring, S. C. (2006). *Situationally embodied curriculum: Relating formalisms and contexts* Retrieved from www.interscience.wiley.com

Blake, C., & Scanlon, E. (2007). Reconsidering simulations in science education at a distance: Features of effective use**.** *Journal of Computer Assisted Learning, 23,* 491–502.

Blasi, L., & Alfonso, B. (2006). Increasing the transfer of simulation technology from R&D into school settings: An approach to evaluation from overarching vision to individual artifact in education. *Simulation & Gaming, 37*(2), 245-267.

Bonanno, P., & Kommers, P. A. M. (2008). Exploring the influence of gender and gaming competence on attitudes towards using instructional games. *British Journal of Educational Technology, 39*(1), 97–109.

Bonk, C., Kim, K., & Zeng, T. (2006). Future directions of blended learning in higher education and workplace learning settings. In C. Bonk & C. Graham (Eds.), *Handbook of blended learning: Global perspectives, local designs* (pp. 550-567). San Francisco, CA: Pfeiffer.

Bos, N. D., Shami, N. S., & Naab, S. (2006). A globalization simulation to teach corporate social responsibility: Design features and analysis of student reasoning, *Simulation & Gaming, 37*(1), 56-72.

Boyes, E. (2007, April 19). *EA:Women "too big an audience to ignore."* Retrieved from http://uk.gamespot.com/pc/strategy/thesims2/news.html?sid=6169357

Bradley, P. (2006). The history of simulation in medical education and possible future directions. *Medical Education, 40*(3), 254-262.

Butler, J. I. (2006). Curriculum constructions of ability: Enhancing learning through teaching games for understanding (TGfU) as a curriculum model, *Sport, Education and Society, 11*(3), 243-258.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3,* 149–210.

Clark, R. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21-29.

Cooper, D. D. (1998). Reading, writing, and reflection. *New Directions for Teaching & Learning, 73*, 47-56.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*, 715-730.

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Cruickshank, D. R., & Telfer, R. (1980). Classroom games and simulations. *Theory Into Practice, 19*(1), 75-80.

Csikszentmihalyi, M. (1997). Finding flow. *Psychology Today, 30*(4), 46-48.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*, 179–202.

De Leng, B. A., Dolmans, D. H. J. M., Muijtjens, A. M. M., & van der Vleuten, C. P. M. (2006). Student perceptions of a virtual learning environment for a problem-based learning undergraduate medical curriculum. *Medical Education, 40*, 568–575.

Dempsey, J. V., Lucassen, B. A., Hynes, L. L., & Casey, M. S. (1996, April). *Instructional applications of computer games.* Paper presented at the 1996 annual meeting of the American Educational Research Association, New York City.

Douglas, A., Miller, B., Kwanza, F., & Cummings, P. (2007). Students' perceptions of the usefulness of a virtual simulation in post-secondary hospitality education, *Journal of Teaching in Travel & Tourism, 7*(3), 1-19.

Eck, R. V. (2006). Digital game-based learning. *Educause Review. 41*(2), 16-30.

Entertainment Software Association. (2013). *2013 essential facts about the computer and video game industry*. Retrieved from http://www.theesa.com/f acts/pdfs/ESA_EF_2013.pdf

Faria, A. J. (1998). Business simulation games: Current usage levels—an update. *Simulation & Gaming, 29*(3), 295-308.

Fong, G. (2006). Adapting COTS games formilitary experimentation, *Simulation & Gaming, 37*(4), 452-465.

Fowler, S. M., & Pusch, M. D. (2010). Intercultural simulation games: A review (of the United States and beyond). *Simulation & Gaming, 41*(1), 94-115.

Garris, R., Ahlers, R., & Driskell, J. (2002). Games, motivation and learning: A research and practice model. *Simulation & Gaming*, *33*(4), 441-467.

Gonzalez, C. S., & Blanco, F. (2008). Integrating an educational 3D game in Moodle. *Simulation & Gaming, 39*(3), 399-413.

Government Accounting Office. (2004). *Military education*: *DOD needs to develop performance goals and metrics for advanced distributed learning in professional military education--Report to the ranking minority member, committee on armed services, house of representatives* [electronic resource]. Washington, DC: Government Accounting Office.

Greenberg, B. S., Sherry, J., Lachlan, K., Lucas, K., & Holmstrom, A. (2008). Orientations to video games among gender and age groups. *Simulation & Gaming 41*(2), 238-259.

Heiny, E. L., & Mundfrom, D. J. (2010). The effectiveness of stepwise discriminant analysis as a post hoc procedure to a significant MANOVA. *Journal of Modern Applied Statistical Methods, 9*(1), 43-51.

Hindle, K. (2002). A grounded theory for teaching entrepreneurship using simulation games. *Simulation & Gaming, 33*(2), 236-241.

Jenson, J., & de Castell, S. (2010). Gender, simulation, and gaming: Research review and redirections. *Simulation & Gaming, 41*(1), 51-71.

Jong, M. S. Y., Shang, J., Lee, F., & Lee, J. H. M. (2008). Harnessing computer games in education, *Journal of Distance Education Technologies, 6*(1), 1-9.

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. San Francisco: Pfeiffer.

Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach.* New York: Springer.

Keller, J. M., & Suzuki, K. (2004). Learner motivation and e-learning design: A multinationally validated process, *Journal of Educational Media, 29*(3), 229-239.

Kim, A. J. (2011, March). *Gamification 101: Designing the player journey*. Retrieved from https://www.youtube.com/watch?v=B0H3ASbnZmc.

Kirkpatrick, D. (1996). Revisiting Kirkpatrick's four-level-model. *Training & Development 50*(1), 54-57.

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology, Research and Development, 42*(2), 7-19.

Krain, M., & Lantis, J. S. (2006). Building knowledge? Evaluating the effectiveness of the global problems summit simulation. *International Studies Perspectives, 7*, 395–407.

Kraska, M. F., & Bentley, T. R. (2004). Perceptions of graduates and their supervisors related to the air and space basic course. *Journal of Industrial Teacher Education, 41*(1), 55-72.

Kruskal, W. H. (1960). Some remarks on wild observations. *Technometrics*, *2*(1), 1-3.

Lantis, J. S. (1998). Simulations and experiential learning in the international relations classroom. *International Negotiation, 3*(1), 39-57.

Laurillard, D. (2001). *Rethinking university teaching: A conversational framework for the effective use of educational technologies*. London: Routledge.

Lepper, M. R., & Chabey, R. W. (1985). Intrinsic motivation and instruction: Conflicting views on the role of motivational processes in computer-based education. *Educational Psychologist, 20*(4), 217-230.

Lohr, L. L. (2008). *Creating graphics for learning and performance; Lessons in visual literacy*. Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.

Lund, A., & Lund, M. (2013). *Laerd statistics.* Retrieved from https://statistics.laerd.com.

Lynch, M. A., & Tunstall, R. J. (2008). When worlds collide: Developing game-design partnerships in universities, *Simulation & Gaming, 39*(3), 379-398.

MacCuish, D. A. (2001). Formative evaluation of the Air Command Staff College distance learning program: A status study. *Education at a Distance, 15*(5), 1-17.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187-198.

McKeachie, W. J. (2002). *McKeachie's teaching tips.* Boston: Houghton Mifflin.

Murray, H. J. R. (1913). *A history of chess*. Northampton, MA: Benjamin Press.

Nelson, W. A., Bueno, K. A., & Huffstutler, S. (1999). If you build it, they will come. But how will they use it? *Journal of Research on Computing in Education, 32*(2), 270-286.

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine.* Cambridge, MA: Perseus Books.

O'Neil, H. F., Wainessa, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal, 16*(4), 455-474.

Pacala, J. T., Boult, D., & Hepburn, K. (2006). Ten years' experience conducting the aging game workshop: Was it worth it? *The American Geriatrics Society 54*, 144–149.

Pedaste, M., & Sarapuu, T. (2006). Developing an effective support system for inquiry learning in a Web-based environment. *Journal of Computer Assisted Learning, 22*, 47–62.

Piaget, J. (1951). *Play, dreams and imitation in childhood.* New York: Norton & Company.

Piaget, J. (1952). *The origins of intelligence in children.* New York: Norton & Company.

Prenski, M. (2001). *Digital game-based learning.* New York: McGraw-Hill.

Procci, K., Lakhmani, S., Hussain, T. S., & Bowers, C. A. (2014). Opening cinematics: Their cost-effectiveness in serious games. *Simulation & Gaming, 45*(1), 93-124.

Reiser, R. A., & Dempsey, J. V. (2007). *Trend and issues in instructional design and technology* (2nd ed.). Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.

Renkl, A., Mandl, H. & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist, 31*(2), 115-121.

Rieber, L. P., & Noah, D. (2008). Games, simulations, and visual metaphors in education: antagonism between enjoyment and learning, *Educational Media International, 45*(2), 77-92.

Rubel, R. C. (2006). Epistemology of war gaming. *Naval War College Review, 59*(2), 108-128.

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. Cambridge, MA: The MIT Press.

Schwalbe, S. R. (1993). War gaming: In need of context. *Simulation & Gaming, 24*, 314-320.

Sedlack, R. E. (2007). Validation of computer simulation training for esophagogastroduodenoscopy: Pilot study. *Journal of Gastroenterology and Hepatology, 22*, 1214–1219.

Serafini, F. W. (2002). Dismantling the factory model of assessment. *Reading & Writing Quarterly, 18*, 67-85.

Smaldino, S. E., Lowther, D. L., & Russell, J. D. (2008). *Instructional technology and media for learning.* Upper Saddle River, NJ: Pearson.

van Ments, M. (1995). The development of simulation and gaming in Britain. *Simulation & Gaming, 26*(2), 224-235.

van Merrienboer, J. J. G., & Kirschner, P. A. (2007). *Ten steps to complex learning: A systematic approach to four-component instructional design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Virvou, M., & Katsionis, G. (2006). On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE. *Computers & Education, 50*, 154–178.

Vreman-de-Olde, C., & de Jong, T. (2005). Scaffolding learners in designing investigation assignments for a computer simulation. *Journal of Computer Assisted Learning 22*, 63–73.

Waters, D. (2006, August 21). *Games industry is "failing women."* Retrieved from http://news.bbc.co.uk/1/hi/technology/5271852.stm

Wideman, H. H., Owston, R. D., Brown, C., Kushniruk, A., Ho, F., & Pitts, K. C. (2007). Unpacking the potential of educational gaming: A new tool for gaming research. *Simulation & Gaming, 38*(1), 10-30.

Wirman, H. (2014). Gender and identity in game-modifying communities. *Simulation & Gaming, 45*(1), 70-92.

Wynne, M. W. (2007). Every airman an ambassador. *Contact, 25*(3), 2.

Zagal, J. P., & Mateas, M. (2010). Time in video games: A survey and analysis. *Simulation & Gaming, 41*(6), 844-868.

APPENDIX A

WEB-BASED SURVEY

This appendix contains a copy of the original survey used for this research. The invitation came from the researchers civilian email account. The invitation paragraph was later modified to include the military email address of the researcher to aid invitees in verifying the authenticity of the invitation email. Additionally question one was modified after nine months to accommodate the inclusion of the VEST simulation responses into the survey database.

Original Invitation

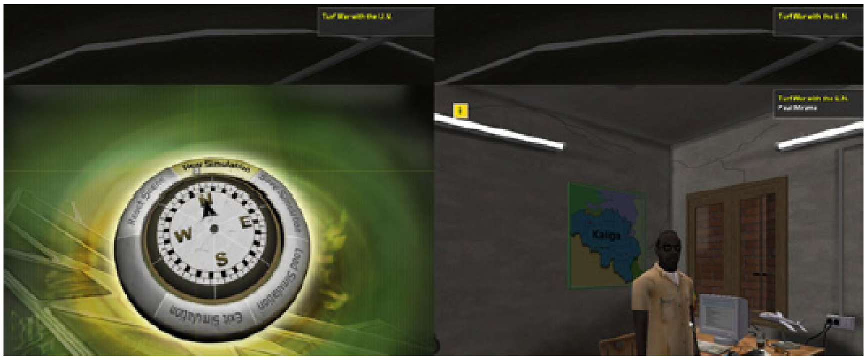
Greetings, I am conducting a study examining the new distance education cultural simulation program at Air War College as part of my PhD studies in educational technology. Air University has approved this research (Survey Control Number 10-107). The purpose of the research is to examine the AWC distance learning cultural simulation program, identify possible areas for improvement, and most importantly create a reference point for other USAF PME programs that might consider creating simulation programs for distance learning courses at their school. All responses will be non-attributable. Your participation in filling out this survey form is strictly voluntary and greatly appreciated. Please answer all survey questions completely and honestly. The survey should take approximately 10 to 15 minutes to complete.

Disclosure Statement:

Participation is voluntary. If you begin participation you may still decide to stop and withdraw at any time. Your decision will be respected and will not result in loss of benefits to which you are otherwise entitled. Returning this survey is acknowledgement that you agree to participate in this survey. If you have any concerns about your selection or treatment as a research participant, please contact the Sponsored Programs and Academic Research Center, Kepner Hall, University of Northern Colorado Greeley, CO 80639; 970-351-1907

Thank you for your time and effort.
Colonel Dennis Armstrong

Original Web-based Survey



Please think back to your expereince in completing the AWC's culture simulation and answer the following questions.

* 1. **Have you completed the cultural simulation program?**

- Yes (Please Continue with the Survey.)
- No (Please discontinue this survey. You have recieved it in error)

2. **How much time did you spend to complete the cultural simulation program?**

- Less than 2 hours
- More than 2 but less than 4 hours
- More than 4 but less than 6 hours
- 6 hours or more

3. **What is your gender?**

- Male
- Female

4. **Which category best describes your age?**

- 30-35
- 36-40
- 41-45
- 46-50
- >50

5. **Which of the following best describes your status?**

- U.S. Air Force
- U.S. Army
- U.S. Navy
- U.S. Marines
- International Officer
- Government Civilian

6. **Which of the following best describes your status?**

- Active Duty
- Reserve
- Guard
- Civilian

7. **Which of the following best describes your occupation?**
   - ○ Rated (flying career field)
   - ○ Non-Rated (non-flying career field)

8. **Which answer best describes your attitude towards video games?**
   - ○ I dislike playing video games.
   - ○ I find video games mildly entertaining.
   - ○ I enjoy playing video games
   - ○ I am very enthusiastic about playing video games.

9. **How often do you currently play video games at home?**
   - ○ Never
   - ○ Seldom
   - ○ Occasionally
   - ○ Routinely

10. **Have you played a role in a wargame or exercise before?**
    ○ Yes ○ No

11. **I feel the knowledge gained from the AWC cultural simulation program has helped prepared me for dealing with cultural issues in future assignments.**

Strongly Disagree 0	1	2	3	4	5	6	Strongly Agree 7
○	○	○	○	○	○	○	○

12. **Learning how to use the simulation was frustrating.**

Strongly Disagree 0	1	2	3	4	5	6	Strongly Agree 7
○	○	○	○	○	○	○	○

13. **Navigating between different parts of the simulation was easy.**

Strongly Disagree 0	1	2	3	4	5	6	Strongly Agree 7
❍	❍	❍	❍	❍	❍	❍	❍

14. **When I first looked at this cultural simulation program, I had the impression that it would be easy for me.**
   - ❍ A. Not true
   - ❍ B. Slightly true
   - ❍ C. Moderately true
   - ❍ D. Mostly true
   - ❍ E. Very true

15. **There was something interesting at the beginning of this cultural simulation program that got my attention.**
   - ❍ A. Not true
   - ❍ B. Slightly true
   - ❍ C. Moderately true
   - ❍ D. Mostly true
   - ❍ E. Very true

16. **This material was more difficult to understand than I would like for it to be.**
   - ❍ A. Not true
   - ❍ B. Slightly true
   - ❍ C. Moderately true
   - ❍ D. Mostly true
   - ❍ E. Very true

17. **After completing the introductory information, I felt confident that I knew what I was supposed to learn from this simulation program.**
   - ❍ A. Not true
   - ❍ B. Slightly true
   - ❍ C. Moderately true
   - ❍ D. Mostly true
   - ❍ E. Very true

18. **Completing the challenges in the simulation gave me a satisfying feeling of accomplishment.**
   - ❍ A. Not true
   - ❍ B. Slightly true
   - ❍ C. Moderately true
   - ❍ D. Mostly true
   - ❍ E. Very true

19. **It is clear to me how the content of this material is related to things I already know.**
   - A. Not true
   - B. Slightly true
   - C. Moderately true
   - D. Mostly true
   - E. Very true

20. **Many of the situations in the simulation had so much information that it was hard to pick out and remember the important points.**
   - A. Not true
   - B. Slightly true
   - C. Moderately true
   - D. Mostly true
   - E. Very true

21. **These materials are eye-catching.**
   - A. Not true
   - B. Slightly true
   - C. Moderately true
   - D. Mostly true
   - E. Very true

22. **There were stories, pictures, or examples that showed me how this material could be important to some people.**
   - A. Not true
   - B. Slightly true
   - C. Moderately true
   - D. Mostly true
   - E. Very true

23. **Completing this cultural simulation program successfully was important to me.**
   - A. Not true
   - B. Slightly true
   - C. Moderately true
   - D. Mostly true
   - E. Very true

24. **The quality of the scenarios helped to hold my attention.**
   - A. Not true
   - B. Slightly true
   - C. Moderately true
   - D. Mostly true
   - E. Very true

25. **This cultural simulation program is so abstract that it was hard to keep my attention on it.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

26. **As I worked on this cultural simulation program, I was confident that I could learn the content.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

27. **I enjoyed this cultural simulation so much that I would like to know more about this topic.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

28. **The scenarios in this simulation look dry and unappealing.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

29. **The content of this material is relevant to my interests.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

30. **The way the information is arranged within the scenarios helped keep my attention.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

31. **There are explanations or examples of how people could use the knowledge in this cultural simulation program.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

32. **The scenarios in this simulation were too difficult.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

33. **This cultural simulation program has things that stimulated my curiosity.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

34. **I really enjoyed working on this simulation.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

35. **The amount of repetition in this cultural simulation caused me to get bored sometimes.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

36. **The content and style of the scenarios in this simulation convey the impression that its content is worth knowing.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

37. **I learned some things that were surprising or unexpected.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

38. **After working on this lesson for awhile, I was confident that I would be able to complete the simulation.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

39. **This simulation was not relevant to my needs because I already knew most of it.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

40. **The wording of feedback after completing the scenarios, or of other comments in this simulation, helped me feel rewarded for my effort.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

41. **The variety of scenarios, helped keep my attention on the simulation program.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

42. **The way scenarios are presented is boring.**
- A. Not true
- B. Slightly true
- C. Moderately true
- D. Mostly true
- E. Very true

43. **I could relate the content of this simulation to things I have seen, done, or thought about in my own life.**
- ❍ A. Not true
- ❍ B. Slightly true
- ❍ C. Moderately true
- ❍ D. Mostly true
- ❍ E. Very true

44. **There is so much information in each scenario that it is irritating.**
- ❍ A. Not true
- ❍ B. Slightly true
- ❍ C. Moderately true
- ❍ D. Mostly true
- ❍ E. Very true

45. **It felt good to successfully complete this simulation.**
- ❍ A. Not true
- ❍ B. Slightly true
- ❍ C. Moderately true
- ❍ D. Mostly true
- ❍ E. Very true

46. **The content of this simulation will be useful to me.**
- ❍ A. Not true
- ❍ B. Slightly true
- ❍ C. Moderately true
- ❍ D. Mostly true
- ❍ E. Very true

47. **I could not really understand quite a bit of the material in this lesson.**
- ❍ A. Not true
- ❍ B. Slightly true
- ❍ C. Moderately true
- ❍ D. Mostly true
- ❍ E. Very true

48. **The good organization of the content helped me be confident that I would learn this material.**
- ❍ A. Not true
- ❍ B. Slightly true
- ❍ C. Moderately true
- ❍ D. Mostly true
- ❍ E. Very true

49. **It was a pleasure to work on such a well-designed simulation.**
   - ❍ A. Not true
   - ❍ B. Slightly true
   - ❍ C. Moderately true
   - ❍ D. Mostly true
   - ❍ E. Very true

50. **What did you like the most about the simulation?**

____
____
____
____

51. **What did you like the least about the simulation?**

____
____
____
____

52. **Are you a volunteer to participate in a follow-up interview if needed?**
   ❍ Yes ❍ No

53. **If yes, please provide an email address or telephone number to contact you:**

Email or ____________________________________

Phone ____________________________________

Thank you for completing this survey. I appreciate your time and effort. An executive summary of the study results can be provided upon request (d.armstrong@msn.com).

APPENDIX B

INSTITUTIONAL REVIEW BOARD APPROVAL

UNIVERSITY of NORTHERN COLORADO
Institutional Review Board (IRB)

May 5, 2010

TO: Gary Heise
School of Sport and Exercise Science

FROM: The Office of Sponsored Programs

RE: Exempt Review of *The Effectiveness, Usability, and Motivational Characteristics of Using Animated Role-Playing Simulations for Military Education,* submitted by Dennis Armstrong (Research Advisor: James Gall)

The above proposal is being submitted to you for exemption review. When approved, return the proposal to Sherry May in the Office of Sponsored Programs.

I recommend approval.

Gary D Heise 19 May 2010
Signature of Co-Chair Date

The above referenced prospectus has been reviewed for compliance with HHS guidelines for ethical principles in human subjects research. The decision of the Institutional Review Board is that the project is exempt from further review.

IT IS THE ADVISOR'S RESPONSIBILITY TO NOTIFY THE STUDENT OF THIS STATUS.

Comments:

25 Kepner Hall ~ Campus Box #143
Greeley, Colorado 80639
Ph: 970.351.1907 ~ Fax: 970.351.1934

APPENDIX C

ONE SAMPLE *T*-TEST ASSUMPTION ANALYSIS

This appendix includes an explanation of the assumptions check for the one sample *T*-test of perceived effectiveness for both the original simulation dataset and the VEST simulation dataset.

T-Test Assumptions

The assumptions for a one-sample *T*-test are that the samples are independent, there are no outliers, and dependent variable is normally distributed (Lund & Lund, 2013). The setup of the online survey ensured every participant could only complete one survey so the data were independent. The other two assumptions were checked individually for each simulation's dataset.

Original Simulation Assumption Check

The original simulation data had no outliers for perceived effectiveness as assessed by an inspection of a boxplot in Figure 3. Perceived effectiveness data for the original simulation were normally distributed as assessed by a visual inspection of the Q-Q plot in Figure 4. The original simulation data met the assumptions for the *T*-test.

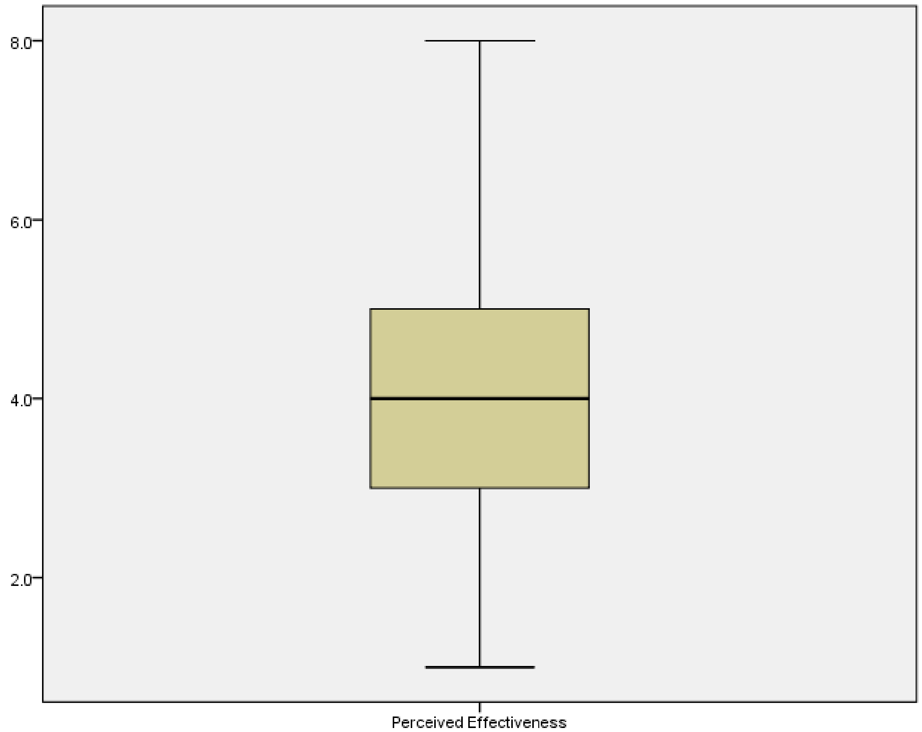


Figure 3. Original simulation boxplot for perceived effectiveness (no outliers).

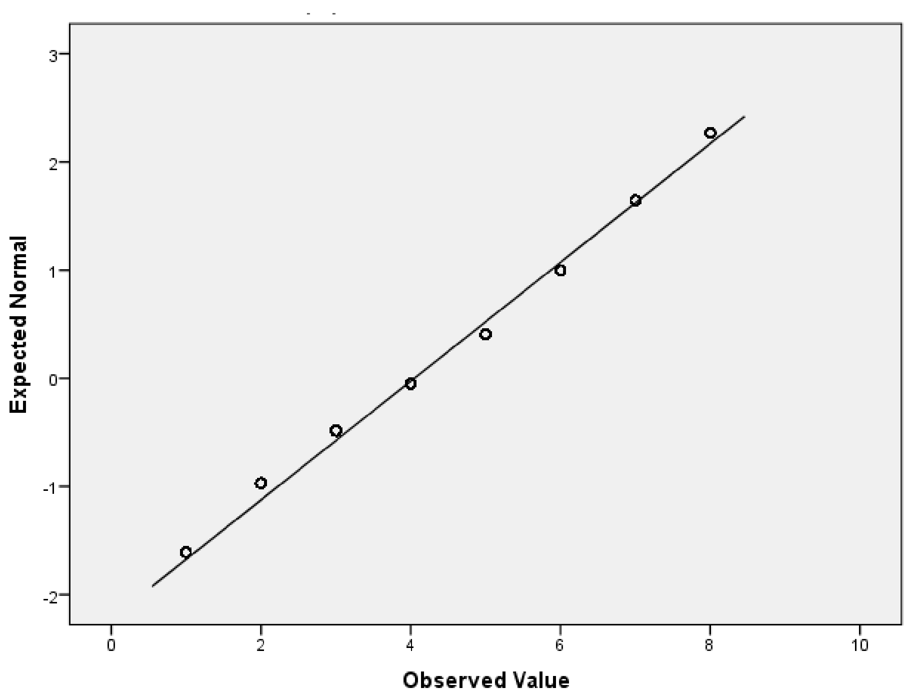


Figure 4. Original simulation Q-Q plot for perceived effectiveness (normal).

Visual Expeditionary Skills Training
Assumption Check

The VEST dataset did have outliers for perceived effectiveness as indicated by the boxplot in Figure 5. Although the boxplot only lists four outliers due to limited space to display the number, there were actually seven outliers. Samples 28, 75, 98, 119, 200, 209, and 265 all were outliers. All seven rated perceived effectiveness at the lowest possible rating of 1. These outliers represented 2.6% of the VEST surveys. All seven samples had negative open-ended comments about the simulation such as "it was a waste of time for someone with my experience," "couldn't understand a single word," and "it kept crashing." There was no reason to suspect data entry errors. For comparison, the original simulation had 96 ratings of 1 out of 898 samples. These represented 10.7% of the original simulation surveys. However, the rating of 1 was not an outlier for the original simulation. The mean of perceived effectiveness was 5.611 for VEST so removing the lower ratings would slightly increase the mean and decrease the variance, which would increase the risk of Type 1 error. Therefore, there should be confidence in a statistically significant result with the outliers present as their inclusion would inhibit rather than promote a significant test result. The outliers were retained in the dataset for the *T*-test analysis. Although the Q-Q plot in Figure 6 showed more deviation from normal than the original simulation dataset, the data were still approximately normal.

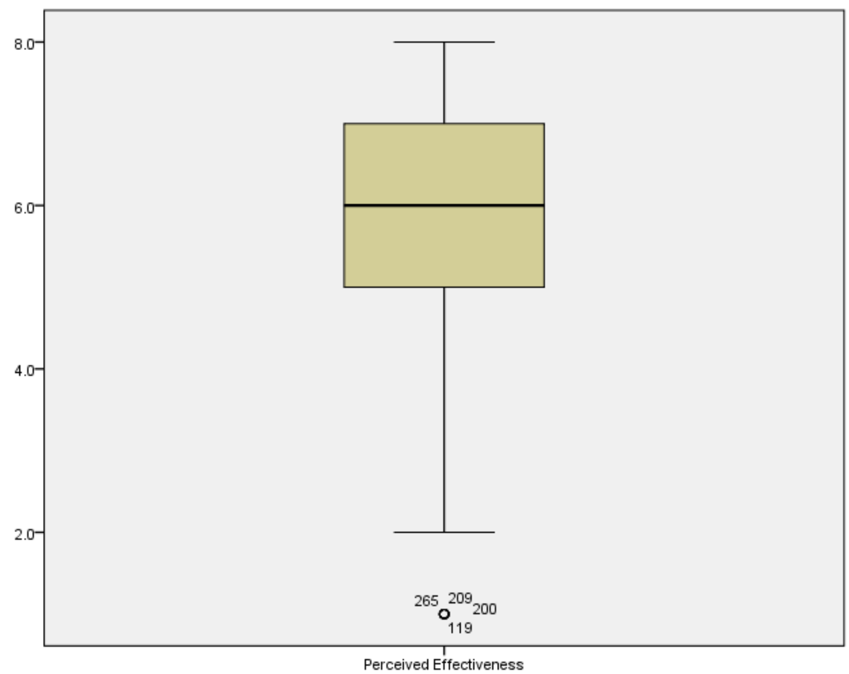


Figure 5. Visual expeditionary skills training simulation boxplot for perceived effectiveness (outliers).

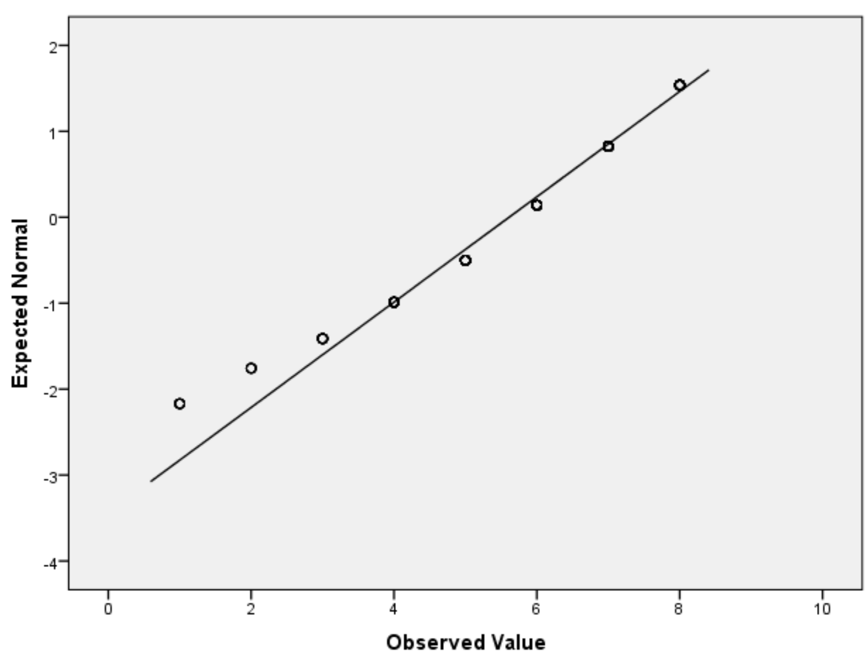


Figure 6 . Visual expeditionary skills training simulation Q-Q Plot for perceived effectiveness (normal).

APPENDIX D

DETAILED FACTOR ANALYSIS

This appendix contains the details of for the factor analysis examining the four aspects of Keller's (2010) motivational model: attention, relevance, confidence, and satisfaction (ARCS).

The factor analysis was accomplished using the combined original simulation (OS) and VEST datasets. The total sample size was 1,163 and each sample contained responses to the 36 questions for Keller's (2010) Instructional Materials Motivation Survey (IMMS) for a total of 41,868 data points. Extraction was accomplished based on using eigenvalues greater than 1 and the varimax method was used for the rotation.

The analysis provided the percent of variance accounted for by component as shown in Table 52. The scree plot is presented in Figure 7. Neither view of the results demonstrated the case for four distinct variables. It appeared there were two main components. A rotated component matrix was used to further investigate the results.

Table 52

Percent of Variance Accounted for by Component

Component	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	12.066	33.517	33.517
2	4.221	11.724	45.241
3	2.211	6.143	51.383
4	2.104	5.845	57.228
5	1.740	4.833	62.061

Note. Extraction Method: Principal Component Analysis.

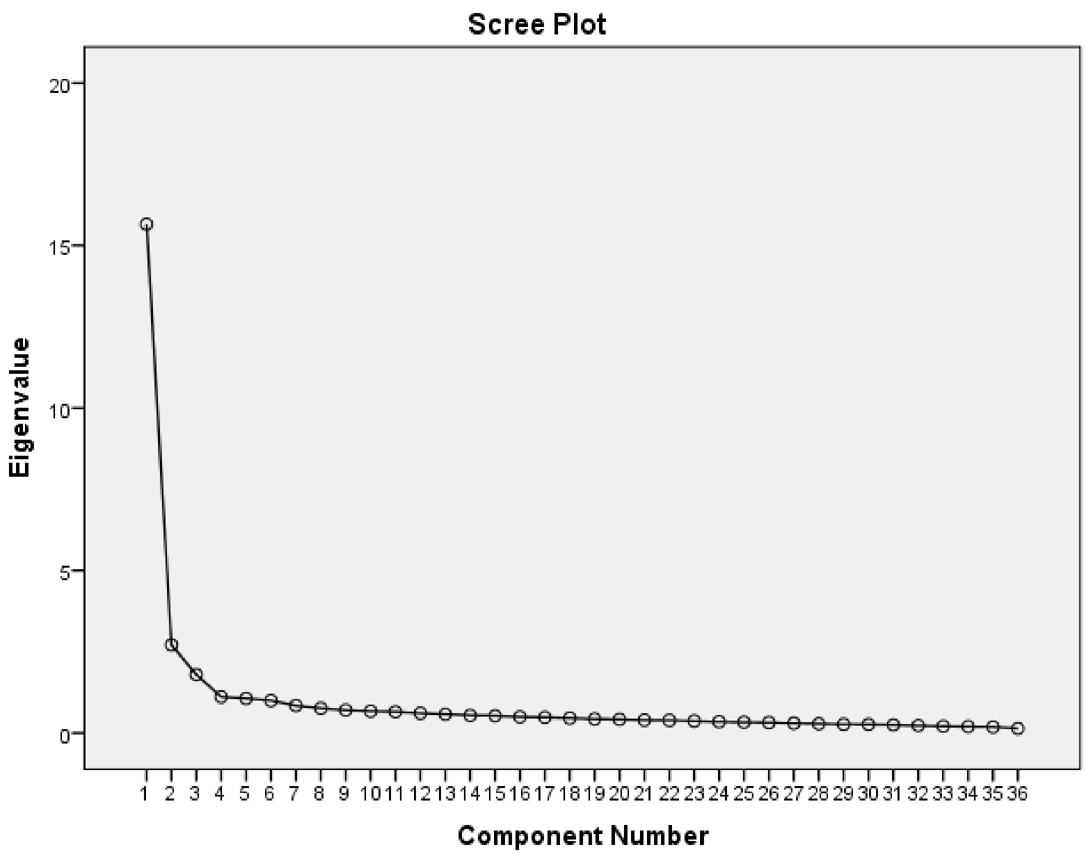


Figure 7. Scree plot for attention, relevance, confidence, and satisfaction component analysis.

The rotated component matrix presented in Table 53 shows the 36 questions grouped by the four IMMS categories of attention, relevance, confidence, and satisfaction. The rotation converged in seven iterations. The factor each variable loaded on the strongest is highlighted. It appeared three variables were largely accounted for with one component. Attention, relevance, and satisfaction all appeared to be loaded the highest on the first factor whereas confidence seemed to be loaded more on the second factor. It appeared confidence was different than the other three but there was no indication within the data that attention, relevance, and satisfaction were clearly distinct from each other.

Table 53

Rotated Component Matrix for Attention, Relevance, Confidence, and Satisfaction Analysis

ARCS Category	Question Number	Component 1	2	3	4	5
Attention	K2	.658	.043	.222	.154	.169
Attention	K8	.689	.111	.313	.089	.125
Attention	K11	.744	.164	.357	.243	.057
Attention	K12	.272	.540	.427	.157	.082
Attention	K15	.416	.296	.669	.128	-.014
Attention	K17	.757	.233	.259	.193	.080
Attention	K20	.823	.129	.200	.122	.014
Attention	K22	.389	.303	.390	.197	-.195
Attention	K24	.763	.085	-.011	.116	-.201
Attention	K28	.757	.189	.289	.204	.056
Attention	K29	.458	.284	.627	.223	-.097
Attention	K31	.192	.666	.322	.024	-.019
Relevance	K6	.514	.223	.070	.187	.357
Relevance	K9	.727	.232	.167	.136	.172
Relevance	K10	.318	.050	.245	.725	.053
Relevance	K16	.604	.049	.199	-.056	.317
Relevance	K18	.652	.297	.006	.030	.185
Relevance	K23	.778	.197	.171	.190	.094
Relevance	K26	.441	.089	.185	.174	-.588
Relevance	K30	.444	.089	.093	-.045	.507
Relevance	K33	.803	.148	.122	.171	.050
Confidence	K1	.033	.068	-.050	.047	.639
Confidence	K3	.131	.718	.110	-.036	.054
Confidence	K4	.396	.276	.116	.221	.384
Confidence	K7	-.013	.677	.178	-.069	.091
Confidence	K13	.461	.483	-.150	.278	.269
Confidence	K19	.200	.744	-.077	.012	-.044
Confidence	K25	.314	.565	-.316	.224	.176
Confidence	K34	.097	.657	.124	.133	.024
Confidence	K35	.751	.281	.034	.166	.038
Satisfaction	K5	.707	.068	.107	.391	-.050
Satisfaction	K14	.798	.120	.020	.053	-.034
Satisfaction	K21	.833	.202	.111	.169	-.025
Satisfaction	K27	.706	.098	.033	.265	-.023
Satisfaction	K32	.342	.055	.052	.766	-.027
Satisfaction	K36	.826	.236	.121	.159	-.006

Note. Keller questions are highlighted by category :attention, relevance, confidence, and satisfaction

This did not negate in any way the extensive research on Keller's (2010) IMMS but this particular case suggested a two factor model be used for analysis since there was no clear distinction in the variance for the variables of attention, relevance, and satisfaction. The three variables of satisfaction, attention, and relevance (SAR) were combined for the two-way MANOVA analysis.

APPENDIX E

DETAILED MULTIVARIATE ANALYSIS OF VARIANCE ASSUMPTION CHECK

This appendix contains a detailed examination of the MANOVA assumptions check for the original simulation dataset using gender and video game experience as the independent variables and perceived effectiveness, usability, confidence, and the combined SAR (satisfaction, attention, and relevance) variable as the dependent variables. The sample size was 898. According to Lund and Lund (2013), there are two requirements and seven assumptions when accomplishing a MANOVA.

The first requirement is the independent variables are categorical, meaning they have two or more categories (Lund & Lund, 2013). The two independent variables were gender and prior video game experience. Gender was captured using the traditional male or female response categories without the use of any additional definitions. Military members are accustomed to filling out paperwork and indicating gender using the traditional criteria of bodily organs. The use of additional categories of "transgender" and "other" was avoided because it would be significantly different from what participants were used to and might create a distraction that could impact the participants' perspectives as they continued with the questionnaire. Additionally, when this research was started, homosexuals could not serve openly in the military. Consequently, adding additional categories would likely provide very little useful information with regard to the research questions because the expected response in the additional categories would likely be very low. Traditional use of gender is a two-category variable by definition. The second independent variable of prior video game experience was based on responses from two Likert-scale questions. As explained previously, these questions were combined and a median split was used to categorize participants as experienced or inexperienced video game users. This resulted is a two-category variable. Therefore, the first requirement was met by both independent variables.

The second requirement was the dependent variables were continuous--measured either as interval or ratio (Lund & Lund, 2013). For each of the dependent variables, perceived effectiveness, usability, effectiveness, confidence, and the combined satisfaction, attention, and relevance variable (SAR) Likert-scale questions were used. As noted in Chapter III on methodology, using Likert-scale responses as an interval scale is debatable but is common practice, and the risk to the analysis is minimal. Therefore, the data met the second requirement.

The first assumption was the data were from independent observations (Lund & Lund, 2013). The surveys were sent via email directly to each individual and the only way to complete the survey was to click on the link provided in the email. The structure of the survey website ensured a specific link from an email could only be used once to complete a survey. After completing the survey, a subsequent selection of the link would thank the participant for already completing the survey. The follow-up email invitation for individuals who had not completed the survey contained the same link so it prevented multiple responses from the same participant. This process helped ensure the first assumption of independent observations was met.

The second assumption was an adequate sample size. While larger sample sizes are better, there must be at least as many samples in each group as there are dependent variables being analyzed (Lund & Lund, 2013). There were six dependent variables and as shown in Table 54, each of the groups was much larger than the minimum so the assumption was met.

Table 54

Sample Size for Original Simulation Groups

	Female	Male
Inexperienced Video Gamers	55	446
Experienced Video Gamers	51	346

N = 898

The third assumption was there were no univariate or multivariate outliers (Lund & Lund, 2013). Like many of the remaining assumptions, this assumption needed to be checked both from univariate and multivariate views. To check for univariate outliers, boxplots of each dependent variable were used for each group of independent variables. The results are presented in Figures 8, 9, 10, and 11, which showed confidence was the only variable that had outliers.

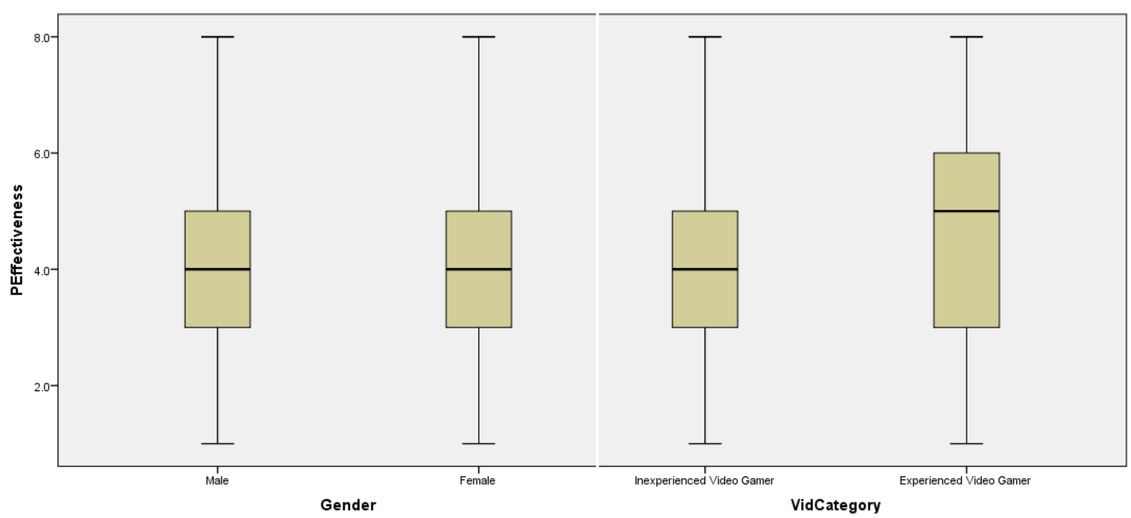


Figure 8. Boxplots of perceived effectiveness by group (*N* = 898).

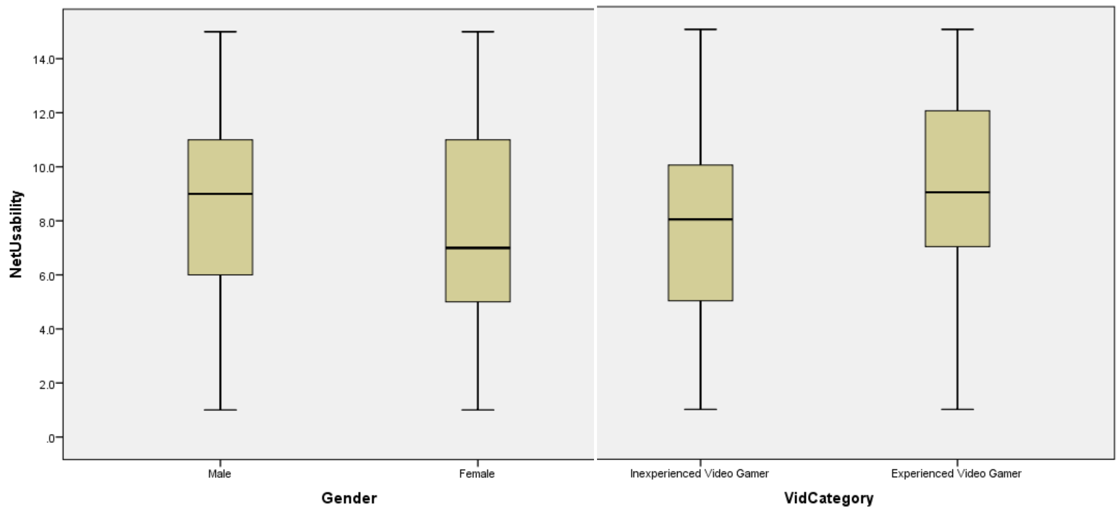


Figure 9. Boxplots of usability by group ($N = 898$).

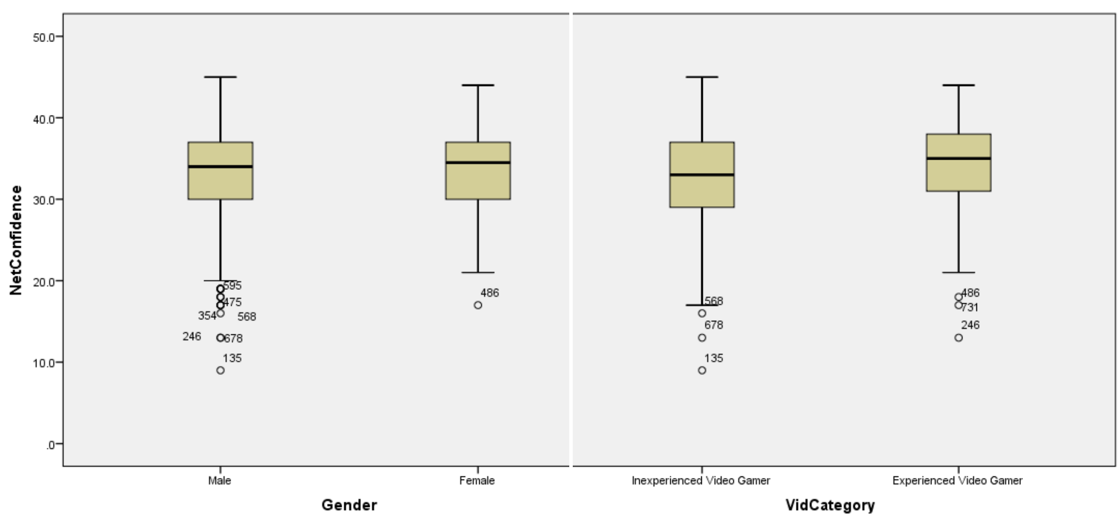


Figure 10. Boxplots of confidence by group ($N = 898$).

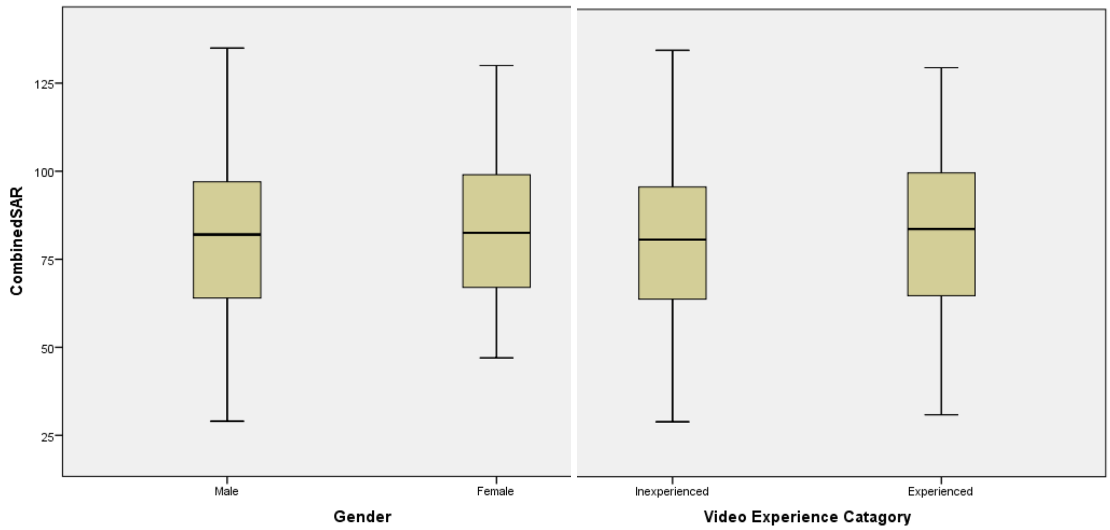


Figure 11. Boxplot of combined satisfaction, attention, and relevance variable by group (N = 898).

There were univariate outliers in all four groups for the confidence variable. However, all the individual outliers from the video game groups were also outliers in the gender groups. The boxplots display some of the identification numbers but due to having multiple samples with the same confidence rating, some of the identification numbers were covered up. The lowest six values for males were outliers. In total, 16 male outliers were represented by those six circles. There was one female outlier, three inexperienced video gamer outliers, and three experienced video gamer outliers. The breakout of outliers is presented in Table 55.

Table 55

Univariate Outliers Broken Out by Group

	Confidence Value						Total
	9	13	16	17	18	19	Samples
Male	1	2	1	4	3	5	16
Female				1			1
Inexperienced*	1	1	1				3
Experienced*		1		1	1		3

*All outliers in the video experience groups are also present in the gender groups.

A close examination of these outliers was important to ensure confidence in the results of the MANOVA. The outliers were inspected and did not appear to be due to entry errors or measurement errors. There were nine confidence questions. The Likert response value range for each question was from 1 to 5. Therefore, the net confidence value could range from 9 to 45. These outliers were all from the low end of the confidence measurement. Most of the written remarks for these outlier samples contained comments consistent with low confidence ratings and included issues like the long time required to complete the simulation, the student not being sure how to get through the simulation, and the use of adjectives such as "hard," "frustrating," and "difficult." These appeared to be genuine data points. Simply removing these points carried risk of altering the results without knowing what effect the outliers had on those results. Kruskal (1960) recommended completing the analysis with the outliers and without the outliers and comparing the results. If the results were similar, then there should be confidence in the results but if they were different, the conclusions would be suspect.

To accomplish the check for multivariate outliers, a Mahalanobis distance was computed using regression procedures. The Mahalanobis distance was compared against the chi-square distance for the degree of freedom equal to the number of dependent variables--in this case, four. The critical value of 18.47 was used. Three cases exceeded this value: sample numbers 253, 246, and 135. While close scrutiny of each outlier was necessary, these were multivariate outliers because of a combination of the dependent variables. This made them more difficult to assess than the univariate outliers. The demographics and numerical values for each of these multivariate outliers are listed in Table 56.

Table 56

Values for Multivariate Outliers

Sample Number	Mahalanobis Number	Gender	Video Game Experience	Perceived Effectiveness	Usability	Confidence	Combined SAR
253	29.4	Male	Exp	8	8	28	50
246	20.7	Male	Exp	5	1	13	49
135	19.1	Male	Inexp	1	1	9	29
		Possible	Range	1 to 8	1 to 15	9 to 45	27 to 135
			Mean	4.05	8.42	33.5	81.6

The furthest multivariate outlier was sample number 253 with a Mahalanobis number of 29.4--well above the 18.47 critical value cut-off. This sample was from a male in the experienced video gamer category and appeared to be an outlier because he rated the simulation the highest possible for perceived effectiveness but below average for all other variables. His wording on the open-ended questions was centered on the simulation being cumbersome to load and often crashing. It is possible he thought the simulation was effective if he looked past the loading/crashing issues but those issues impacted his motivation. This sample could not be dismissed as a simple entry error.

The next furthest multivariate outlier was sample number 246 with a Mahalanobis number of 20.7. This sample was a male categorized as an experienced video gamer. He gave a higher than average rating for perceived effectiveness but the lowest possible rating for usability and low ratings for confidence and the combined SAR variable. His remarks mentioned the hints in the simulation were not helpful. It could be he considered the simulation effective but difficult to use. Like the previous multivariate outlier, this did not appear to be a data entry issue.

The closest multivariate outlier to the critical value was sample number 135 with a Mahalanobis number of 19.1, which was also above the critical value of 18.47. This sample was also a male categorized as inexperienced with video games. This sample had the lowest possible ratings for perceived effectiveness, usability, and confidence. It also had a very low rating for the combined SAR variable. For the open-ended question about what he liked the most about the simulation, he answered "NOTHING" and for the open-ended question about what he liked the least his answer included "This exercise was a huge source of frustration and had no relevance." The ratings were consistent with these remarks so it did not appear to be a data entry error.

Even though three multivariate outliers out of 898 samples was a low percentage and the MANOVAs were somewhat robust to outliers, the same solution identified for the univariate outliers was used. Two of the multivariate outliers were also univariate outliers so the total number of outliers was 18. The outliers accounted for just 2% of the 898 samples. A MANOVA was conducted with all 898 samples and a second MANOVA was conducted with the outliers removed ($N = 880$) to see if there was any difference.

The fourth assumption was the dependent variables were normal. This assumption pertained to both normality of each individual dependent variable and multivariate normality (Lund & Lund, 2013). If the variables had multivariate normality, they would have univariate normality. Therefore, the check focused on assessing multivariate normality, which was done by checking the normality of each of the four groups for each dependent variable. Although this method was not an exact check of multivariate normality, it provided a reasonable assessment of it (Lund & Lund, 2013).

The assumption of normality was visually examined using Q-Q plots for each variable by group. The Q-Q plots for perceived effectiveness presented in Figure 12 show an "S" shape indicative of kurtosis but none of the plots appeared to be greatly different than normal.

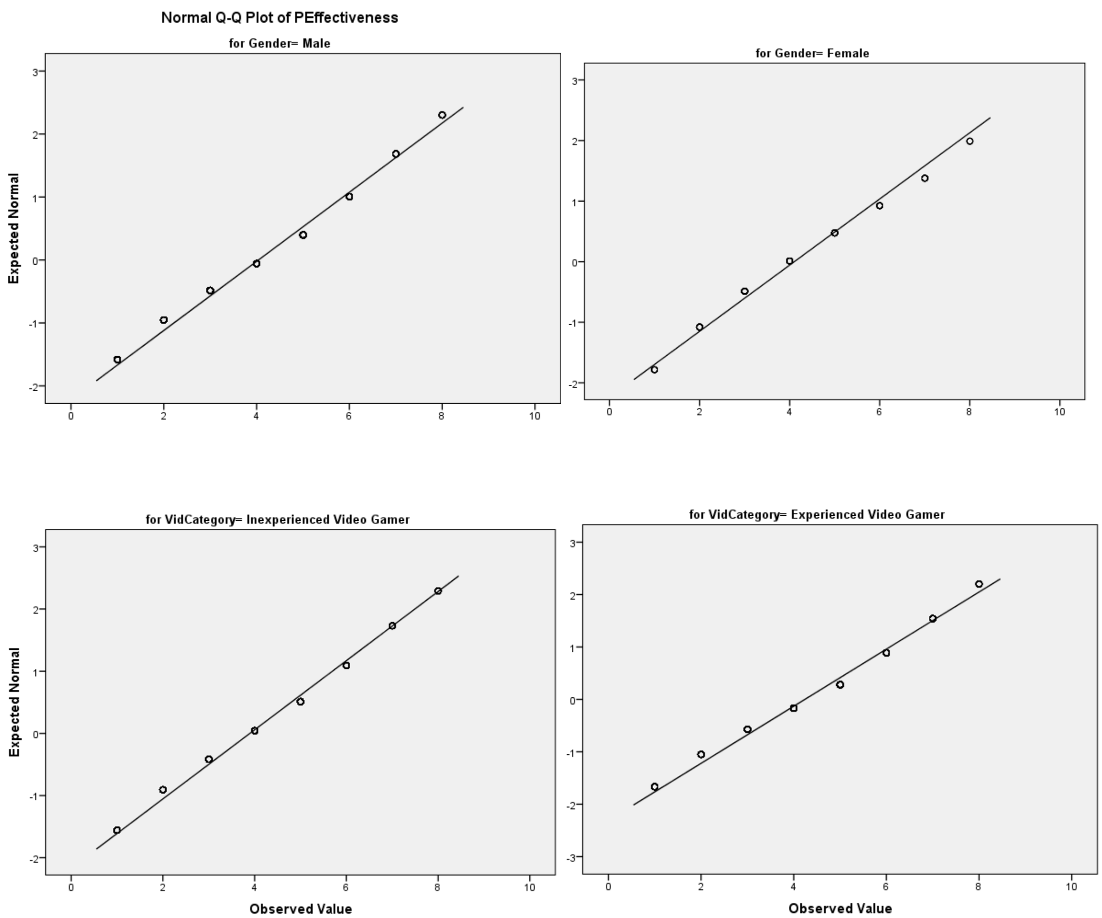


Figure 12. Q-Q plots of perceived effectiveness by group (N = 898).

The Q-Q plots in Figure 13 for usability appeared to have slightly greater kurtosis than the perceived effectiveness plots but still appeared to be nearly normal.

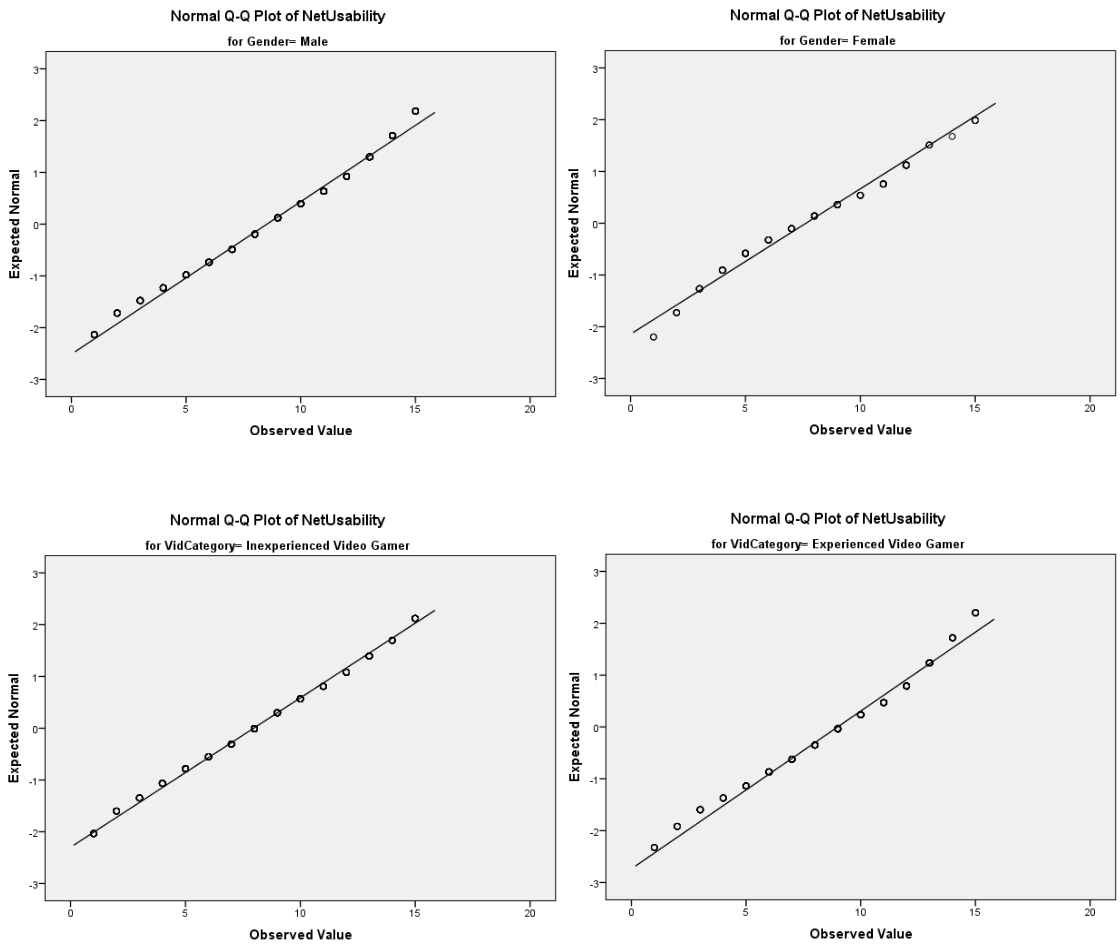


Figure 13. Q-Q plots of usability by group (*N* = 898).

The Q-Q plots shown in Figure 14 for confidence showed a left skew to the data. Overall, they still appeared to be "nearly" normal.

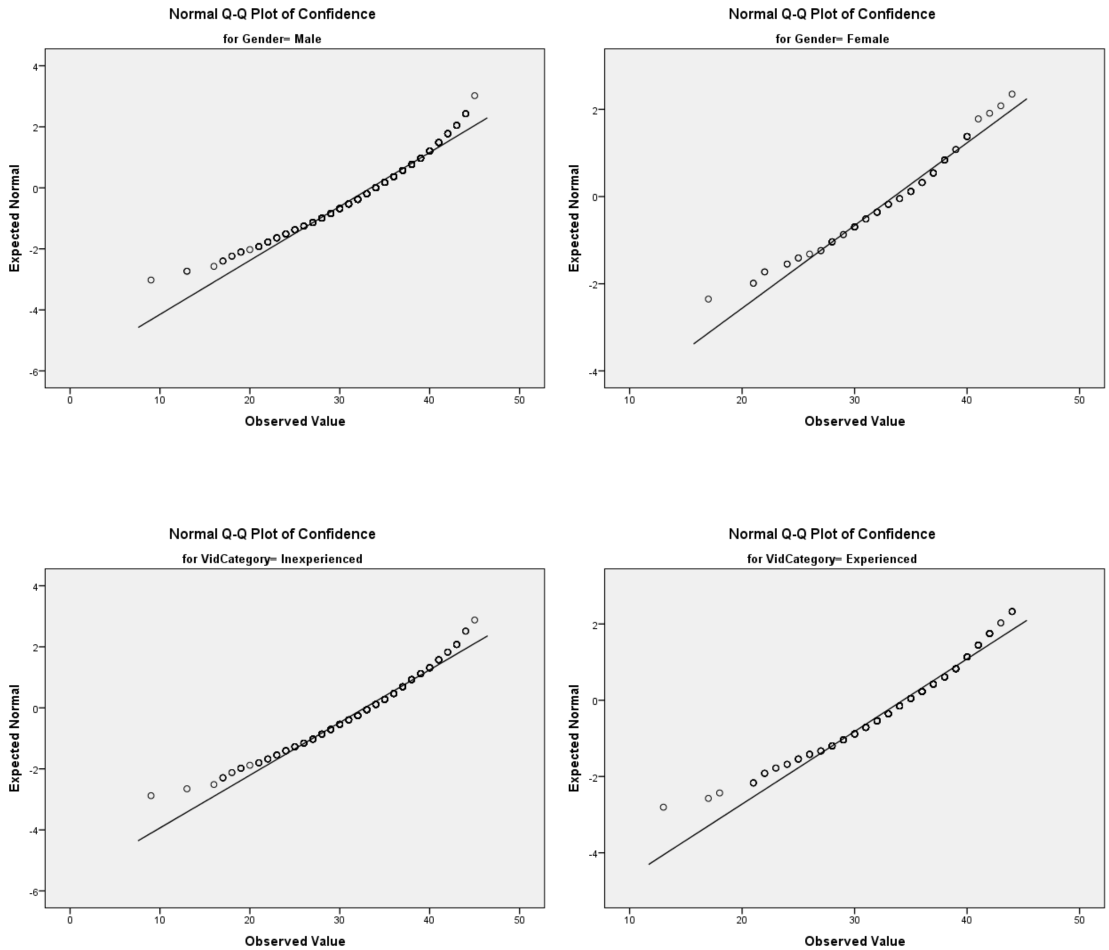


Figure 14. Q-Q plots of confidence by group (*N* = 898).

The Q-Q plots for the combined SAR (satisfaction, attention, relevance) variable shown in Figure 15 all showed some kurtosis. However, they still appeared approximately normal.

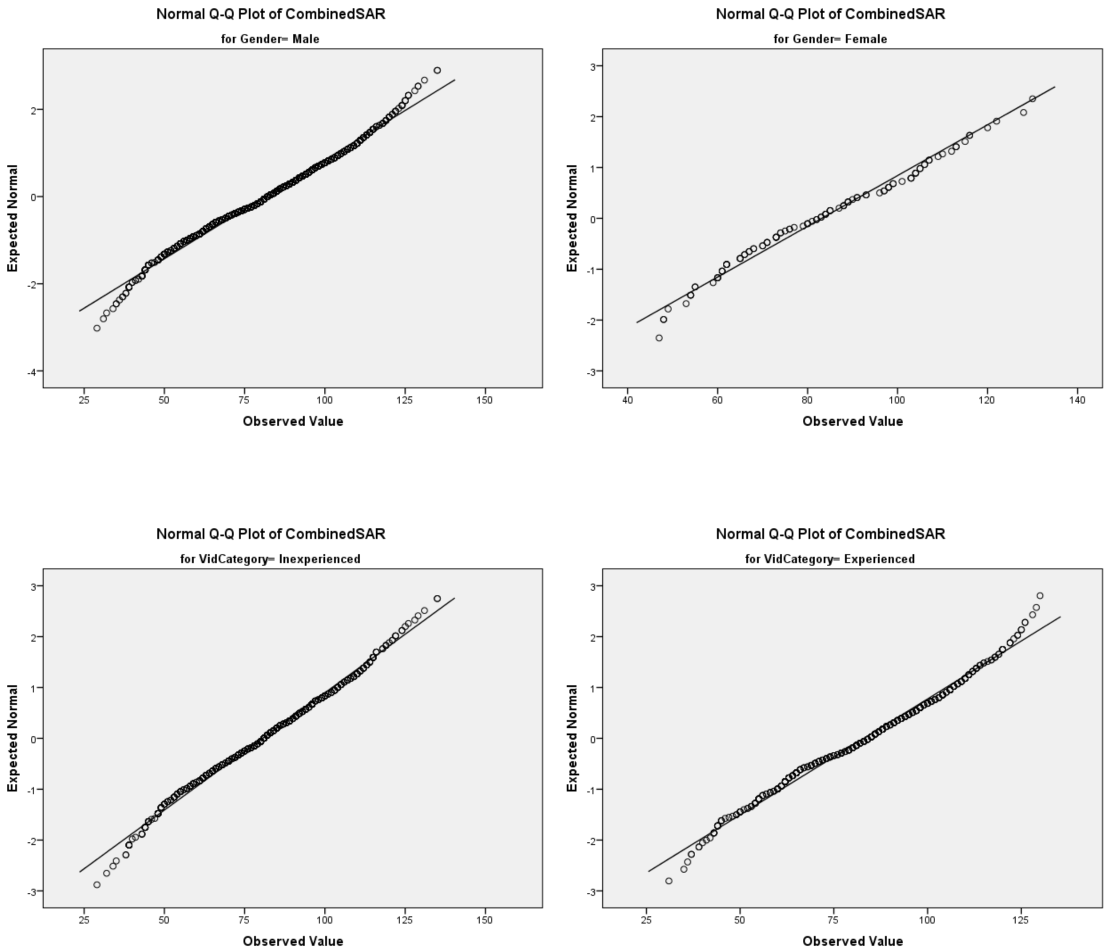


Figure 15. Q-Q plots of combined **s**atisfaction, attention, and relevance by group (*N* = 898).

The visual appearance of the normality curves indicated while there were small departures from normal, the data could be considered nearly normal.

The fifth assumption was linearity, which looks at the relationships between the dependent variables (Lund & Lund, 2013). The scatterplot matrixes shown in Figures 16, 17, 18, and 19 were used to check this assumption. Visually, the scatterplots indicated a linear relationship existed for each of the pairs; therefore, the assumption of linearity was met.

The visual inspection indicated the most linearity appeared between perceived effectiveness and the combined SAR variable. The least linearity appeared between perceived effectiveness and usability.

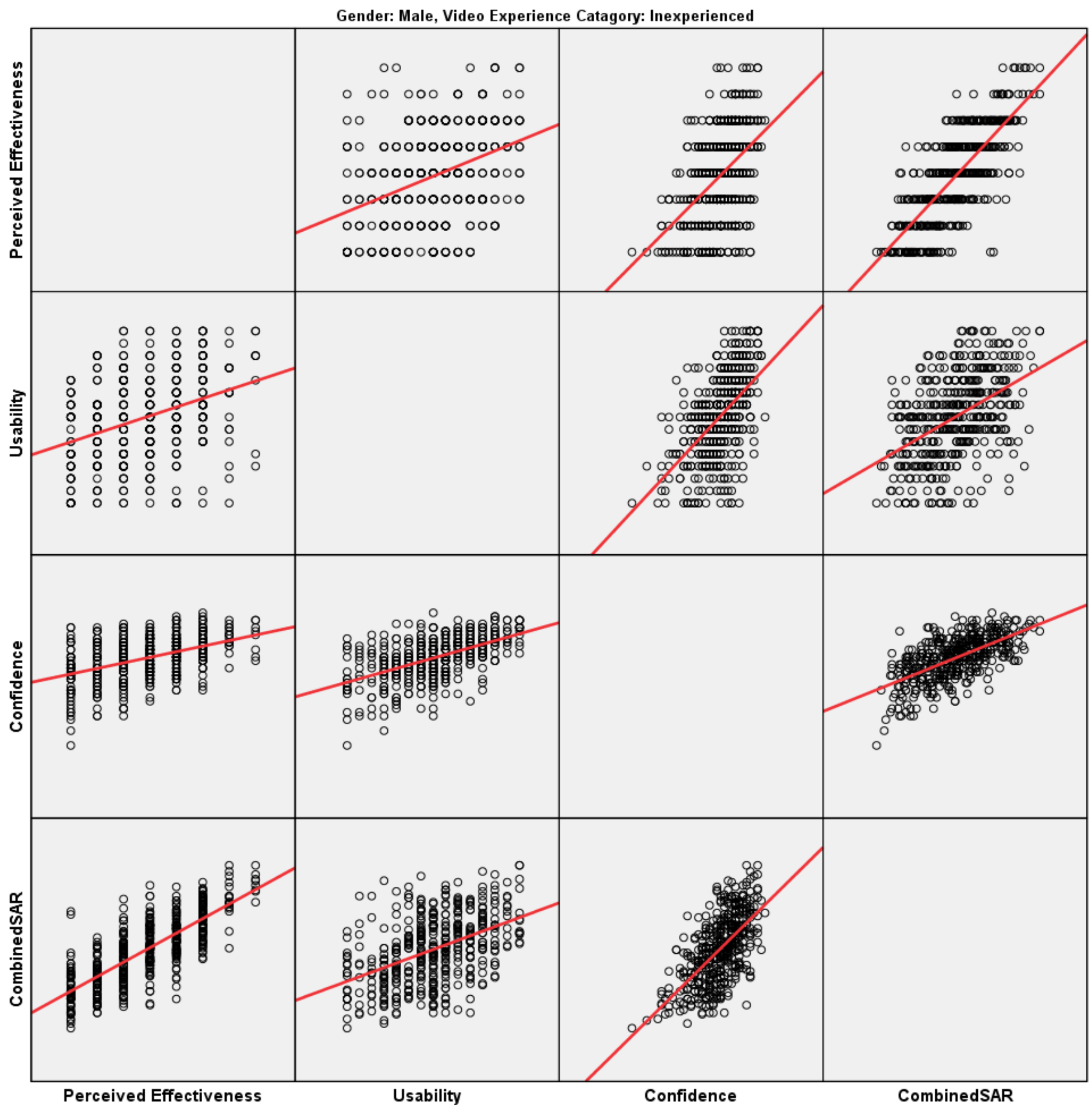


Figure 16. Scatterplot male inexperienced video gamers ($N = 898$).

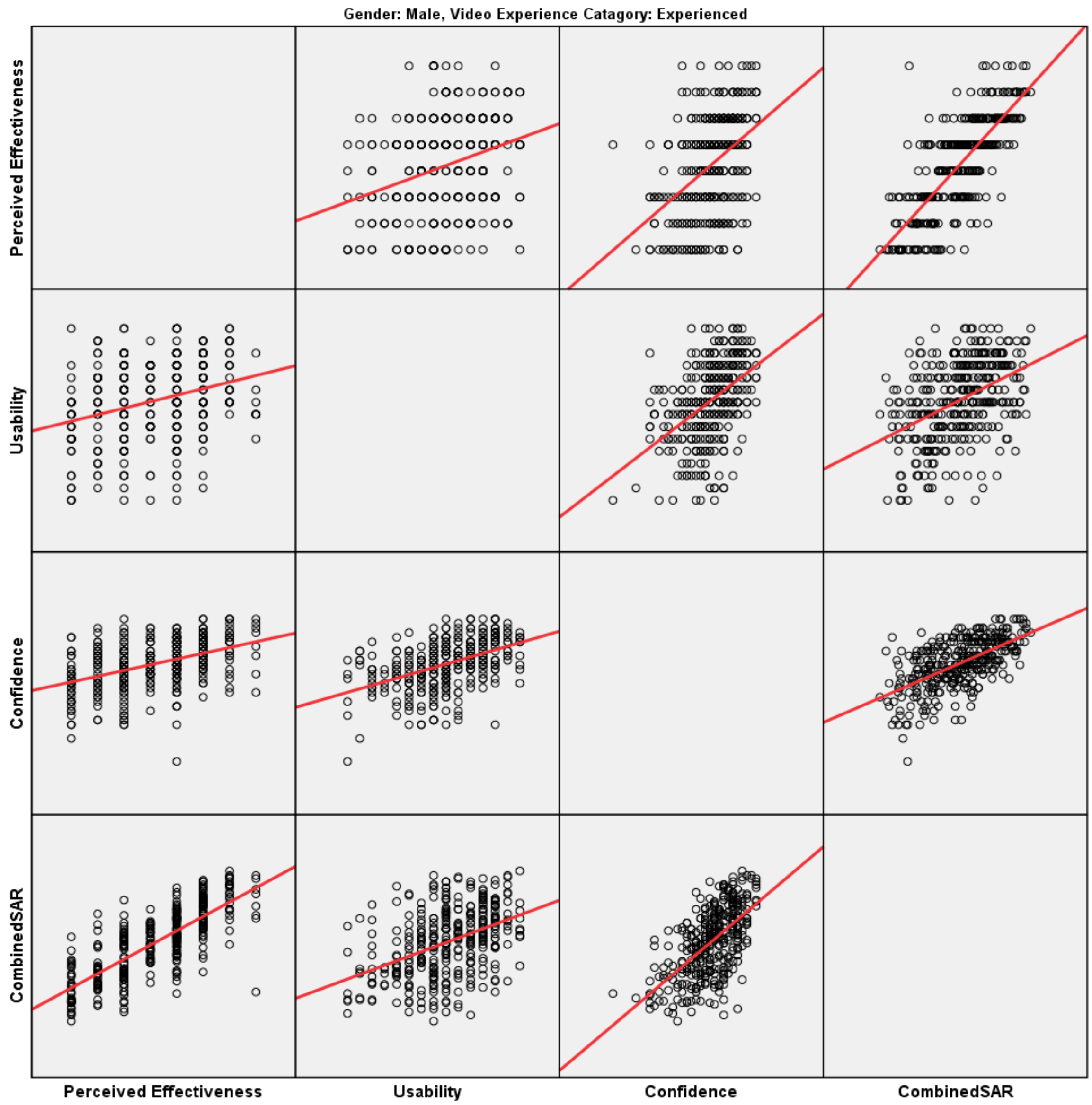


Figure 17. Scatterplot male experienced video gamers (N =898).

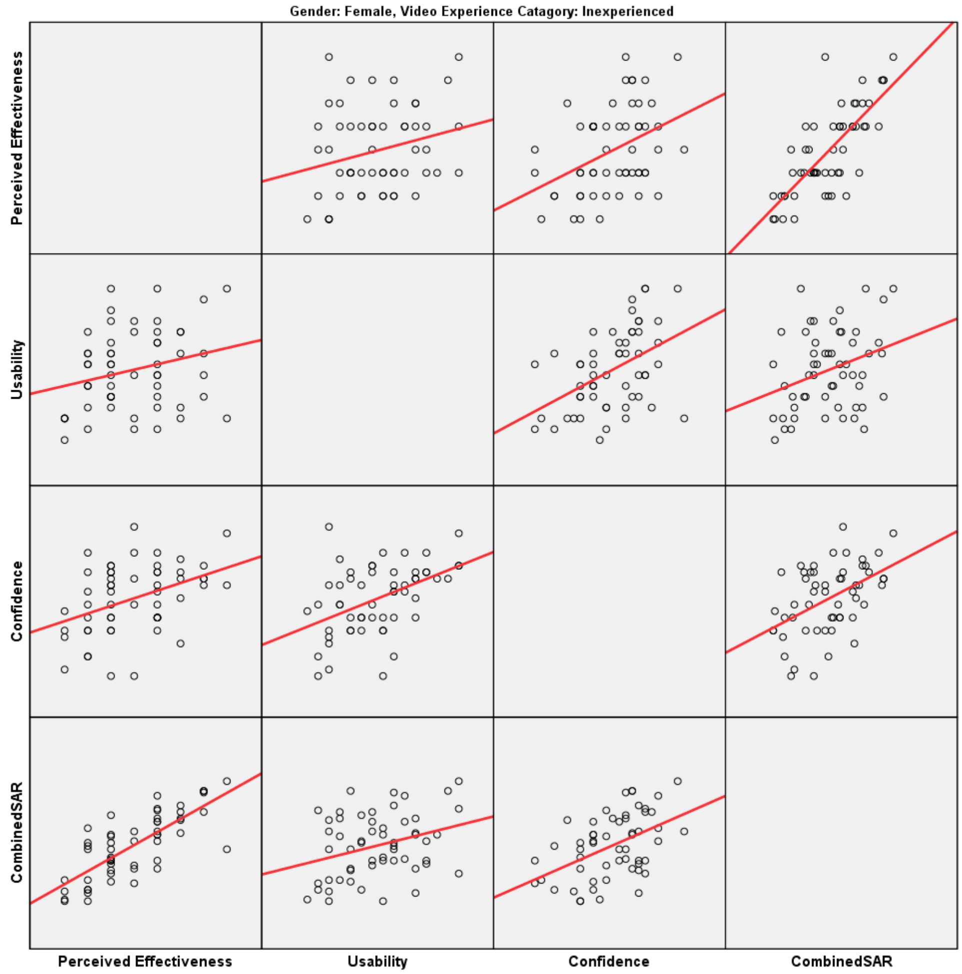


Figure 18. Scatterplot female inexperienced video gamers ($N = 898$).

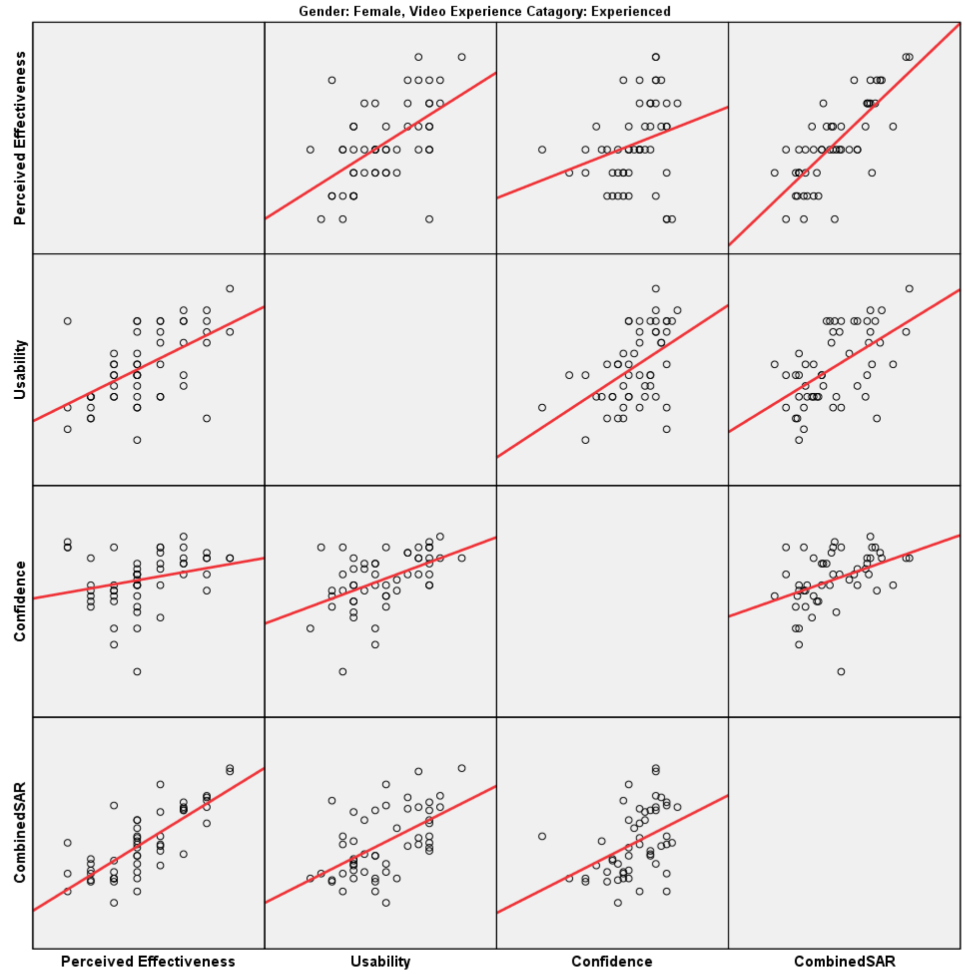


Figure 19. Scatterplot female experienced video gamers (N = 898).

The sixth assumption was the existence of the homogeneity of variance and co-variances. This was tested using both the Levene's test for individual dependent variables and Box's M for the covariance. Both tests were designed so if significance was found (in this case, less than 0.05), it indicated the sample did not meet the assumption and the variances needed to be treated as unequal. Values of greater than .05 meant the variables met the assumption. Values from the Levene's test for each dependent variable are listed in Table 57. All four variables met the requirements for assuming homogeneity of variances. The check for homogeneity of variances within groups (covariance) was checked using Box's M, which had a *p* value of 0.617, so the null hypothesis was rejected and homogeneity of co-variances existed across the groups.

Table 57

Levene's Test of Variances for Each Dependent Variable

Variable	Levene's Test Original Simulation
Perceived Effectiveness	0.609
Usability	0.304
Confidence	0.284
SAR combined	0.675

The seventh and final assumption was no multicollinearity. The dependent variables should be moderately correlated. If they are too low, there is no reason to do the MANOVA and if they are too high (>0.90), it would be indicate multicollinearity, which would be problematic for running a MANOVA (Lund & Lund, 2013). Multicollinearity was checked using Pearson's correlation coefficients between each of the dependent variables. The results are listed in Table 58. All the coefficients indicated moderate correlations with a low of 0.353 for perceived effectiveness and usability and a high of 0.774 for perceived effectiveness and the combined SAR variable. All pairs showed statistical significance using an alpha of 0.01. Overall, the data met the assumption of no multicollinearity.

Table 58

Multicollinearity Check Using Pearson's Coefficients

		Perceived Effectiveness	Usability	Confidence	Combined SAR
Perceived Effectiveness	Pearson Correlation	1	.353**	.445**	.774**
	Sig. (2-tailed)		.000	.000	.000
	N	898	898	898	898
Usability	Pearson Correlation	.353**	1	.523**	.448**
	Sig. (2-tailed)	.000		.000	.000
	N	898	898	898	898
Confidence	Pearson Correlation	.445**	.523**	1	.605**
	Sig. (2-tailed)	.000	.000		.000
	N	898	898	898	898
Combined SAR	Pearson Correlation	.774**	.448**	.605**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	898	898	898	898

** Statistical significance using an alpha of 0.01.

Except for outliers, all MANOVA requirements and assumptions were met with this dataset as shown in Table 59. Seventeen univariate outliers existed for the confidence variable. There were three multivariate outliers but two of those were also in the group of univariate outliers. Therefore, there were 18 outliers total. The outliers seemed to be valid numbers and 18 outliers out of 898 samples was only 2% of the surveys. However, to ensure the outliers did not have a significant effect on the results, two MANOVAs were run--one with the outliers present and one with the outliers removed—to test the effect the outliers had on the results.

Table 59

Results of Requirements and Assumption Check

MANOVA Requirements and Assumptions	Univariate	Multivariate
R1. DV measured as interval	✓	✓
R2. IV are categorical	✓	✓
A1. Independent Observations	✓	✓
A2. Adequate Sample Size	✓	✓
A3. No Outliers	17 identified	3 identified
A4. Normality	✓	✓
A5. Linearity	✓	✓
A6. Homogeneity of Variance	✓	✓
A7. Multicollinearity	✓	✓

APPENDIX F

MULTIVARIATE ANALYSIS OF VARIANCE ASSUMPTION CHECK FOR VISUAL EXPEDITIONARY SKILLS TRAINING

This appendix contains a detailed examination of the MANOVA assumptions check for the VEST dataset using gender and video game experience as the independent variables and perceived effectiveness, usability, confidence, and the combined SAR (satisfaction, attention, and relevance) variable as the dependent variables. The sample size was 265 for the VEST dataset. According to Lund and Lund (2013), there are two requirements and seven assumptions when accomplishing a MANOVA.

The two requirements and first assumption were met as described in Appendix E. Although this was a different dataset, explanation of these three remained the same; the independent variables were still categorical, the dependent variables were continuous, and the observations were independent.

The second assumption that the sample size was adequate was also met. The smallest group was 14, which was greater than the number of dependent variables so the assumption was met.

The third assumption was no univariate or multivariate outliers existed. A check for univariate outliers was accomplished using boxplots of each dependent variable for each group of independent variables. Seventeen surveys contained outliers. Some surveys had outliers for multiple variables for a total of 25 outliers for the four variables. There was no indication these outliers resulted from data entry issues. As a result, a second dataset with outliers removed was created similar to the analysis for the original simulation. If the adjusted dataset had found significance, the original dataset would have been analyzed with separate MANOVA tests to determine how much impact the outliers had on the adjusted results. The following paragraphs discuss the outliers for each variable in detail with information about comments related to the ratings.

Perceived effectiveness had seven outliers. All seven were males who rated perceived effectiveness as a 1, the lowest possible rating (see Figure 20). The lowest non-outlier rating was 2. The comments focused on two issues. First, the simulation ran extremely slow due to bandwidth issues (three cases) and second, some people with prior experience in dealing with cultures felt they should not have to accomplish the simulation (four cases). These did not appear to be data entry problems. The bandwidth issue was about connectivity external to the simulation itself but it clearly impacted the simulation experience. Concerns about people who have had prior experience with culture not being required to accomplish the simulation should be considered for AWC curriculum planning purposes. However, this issue did not reflect whether the simulation was actually effective in its design and operation but rather the appropriateness of the material for those students. Additionally, the simulation was not designed for students with significant experience in dealing with cultures and, subsequently, would not be very effective for those individuals.

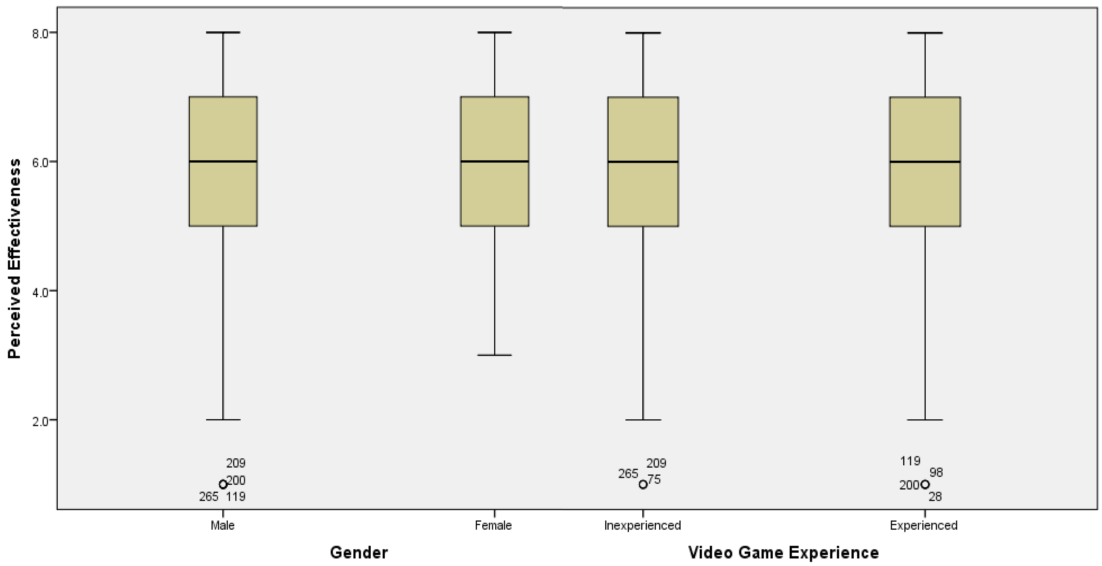


Figure 20. Boxplots of perceived effectiveness by group for visual expeditionary skills training ($N = 265$).

The usability variable had seven outliers as shown in Figure 21. The inexperienced group only had one outlier, which received a rating of 1. The survey range for usability was from 1 to 15. The comment identified a bandwidth issue that created problems playing the simulation. The experienced group had six outliers. One survey had no comments, five surveys mentioned bandwidth issues, and one survey mentioned navigation issues but did not specify if bandwidth was a factor in having trouble navigating. The ratings for these outliers ranged from 1 to 5. There did not appear to be any inconsistency between the comments and the ratings.

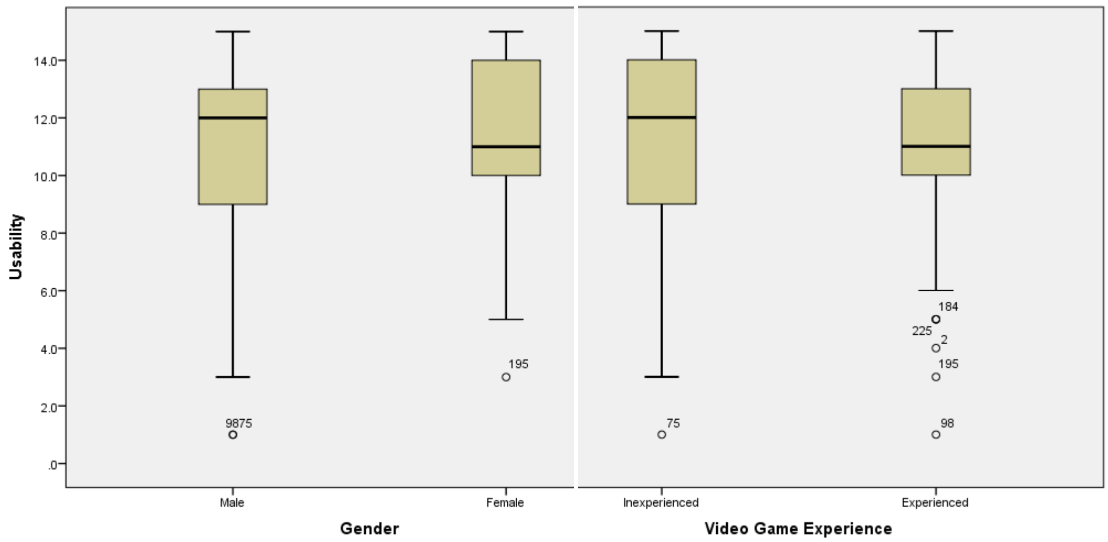


Figure 21. Boxplots of usability by group for visual expeditionary skills training (*N* = 265).

There were eight outliers for the confidence variable as shown in Figure 22. Two comments mentioned bandwidth issues, two implied they had expertise already, and the rest were not specific for confidence. There is no reason to think these were input errors. The ratings ranged from 24 to 31.

Figure 22. Boxplots of confidence by group for visual expeditionary skills training (*N* = 265).

There were three outliers for the combined SAR variable as shown in Figure 23. All three were samples identified as outliers for at least one other dependent variable. Comments for all three cited download speed issues. The ratings seemed consistent with those comments.

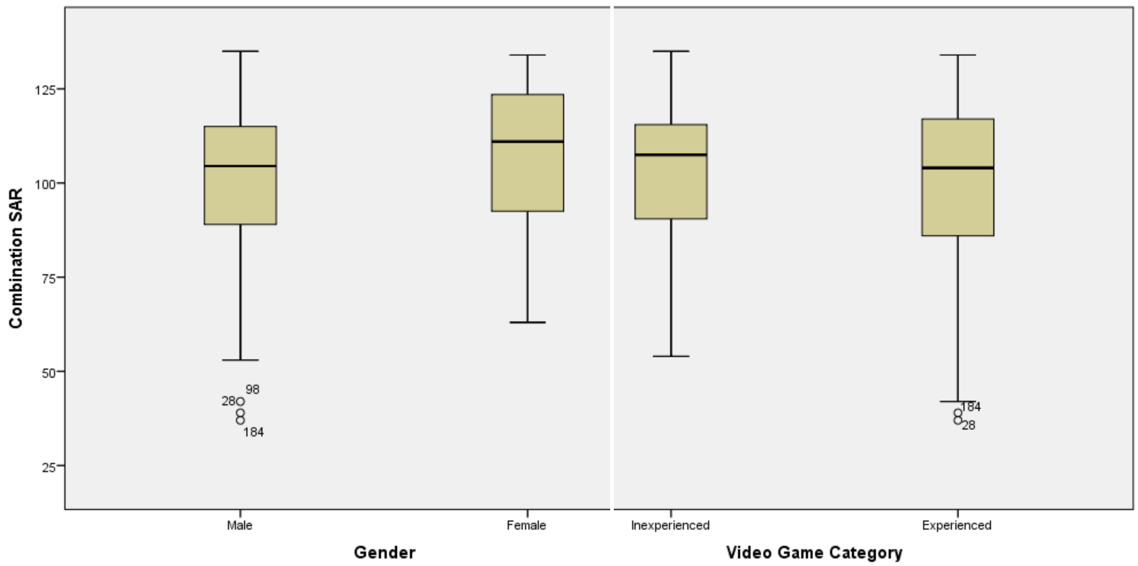


Figure 23. Boxplots of combined satisfaction, attention, and relevance by group for visual expeditionary skills training ($N = 265$).

There were 17 univariate outliers for the 898 samples from original simulation data, which was approximately 1.9% of the surveys compared to the 17 outliers for the 265 samples from the VEST data--approximately 6.4% of the surveys. Additionally, outliers were present in several of the variables for the VEST dataset compared to outliers only being present for confidence for the original simulation. This was consistent with the smaller standard deviations for the VEST data. The data indicated a tighter grouping of scores toward the upper part of the scale compared to the original simulation, causing scores that would not have been outliers for the original simulation to be outliers for the VEST data.

To check for multivariate outliers, a Mahalanobis distance was computed for each sample. Three samples were above the critical value of 18.47, indicating three multivariate outliers as shown in Table 60. Sample number 184 had the largest Mahalanobis distance of 28.4. This sample had a higher than average rating for perceived effectiveness, low ratings for usability and confidence, and a very low rating for the combined SAR variable. The comments noted the participant already knew the information, there were issues with download speed, and he had also tried the original simulation, which he found to be painful and frustrating. The higher rating for perceived effectiveness was surprising. However, even though he knew the information and the simulation ran poorly, it is possible he thought if download speeds were not an issue, the simulation presented the material effectively. Sample 98 had the second largest Mahalanobis number with a distance of 23.5. This sample had very low ratings for all four variables. The comments described download speed problems preventing the simulation from working properly. The comments were consistent with the ratings. The final multivariate outlier was sample 75 with a distance of 19.1. He rated perceived effectiveness and usability at the lowest possible ratings, confidence was slightly below the average, and the combined SAR variable was below average. Again, the comments noted download speed issues. The comments were consistent with the ratings. All three outliers appeared to be valid ratings. Additionally, all three outliers were samples that were also univariate outliers. Therefore, the total number of outliers in the dataset was 17. The comparative dataset with the outliers removed had a sample size of 248.

Table 60

Values for Multivariate Outliers for Visual Expeditionary Skills Training

Sample Number	Mahalanobis Number	Gender	Video Game Experience	Perceived Effectiveness	Usability	Confidence	Combined SAR
184	28.4	Male	Exp	6	5	31	37
98	23.5	Male	Exp	1	1	24	42
75	19.1	Male	Inexp	1	1	37	76
		Possible Range		1 to 8	1 to 15	9 to 45	27 to 135
		Mean for VEST		5.6	11.1	39.5	102.0

N = 265

The fourth assumption was the dependent variables were univariately and multivariately normal. Q-Q plots were created for each variable by group. The Q-Q plot for perceived effectiveness in Figure 24 had a slight "S" shape indicative of kurtosis but it did not appear to be significantly different than normal.

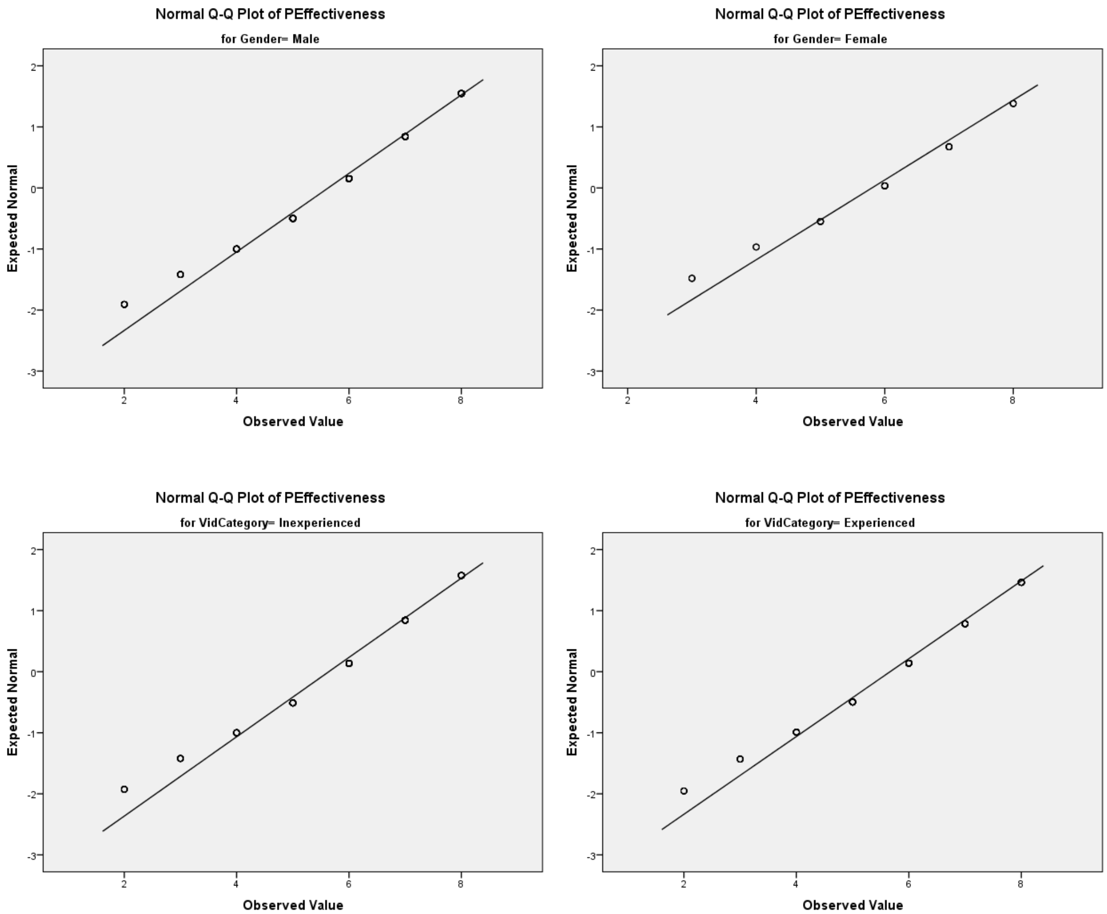


Figure 24. Q-Q plots of perceived effectiveness by group for visual expeditionary skills training ($N = 265$).

The Q-Q plots in Figure 25 for usability appeared to have slightly more kurtosis than the perceived effectiveness plots but were still regarded as nearly normal.

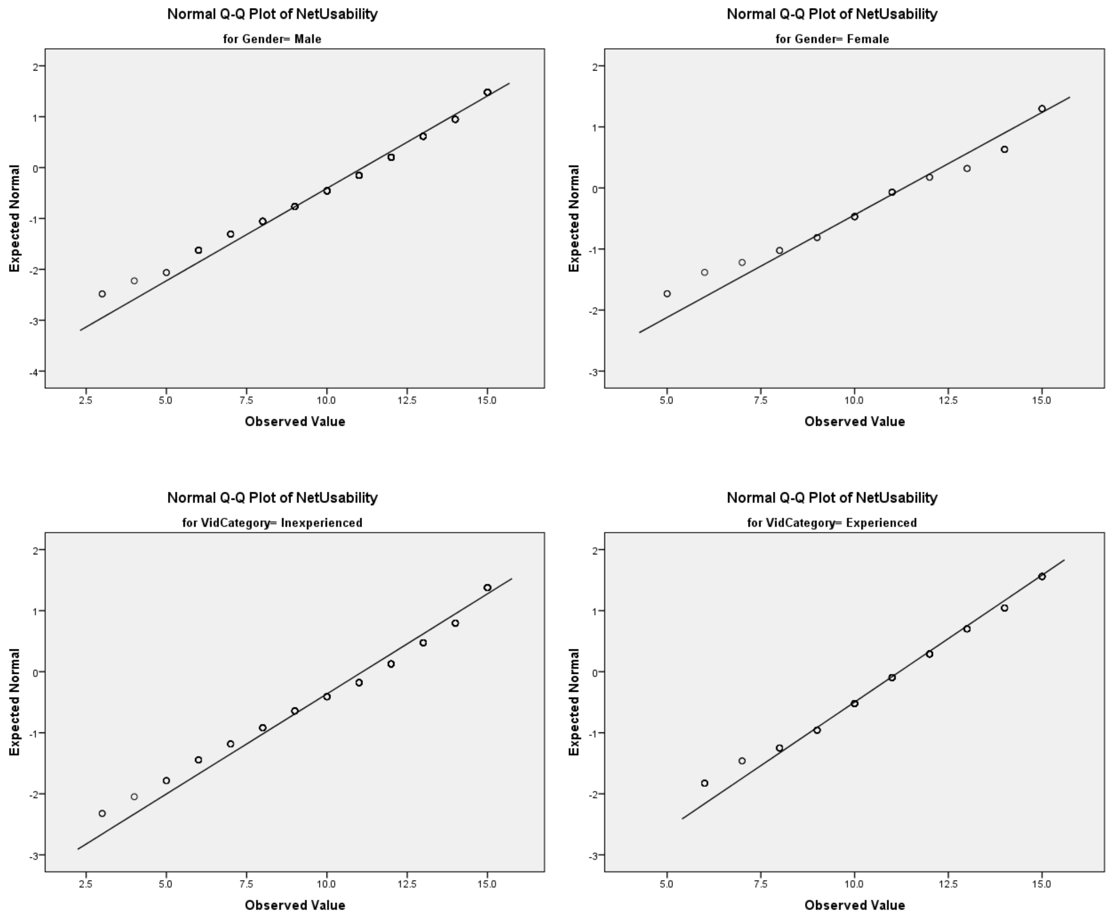


Figure 25. Q-Q plots of usability by group for visual expeditionary skills training (*N* = 265).

The Q-Q plots for confidence as shown in Figure 26 had a pronounced skewing to the left. These plots indicated some deviation from normal, especially in the lower range.

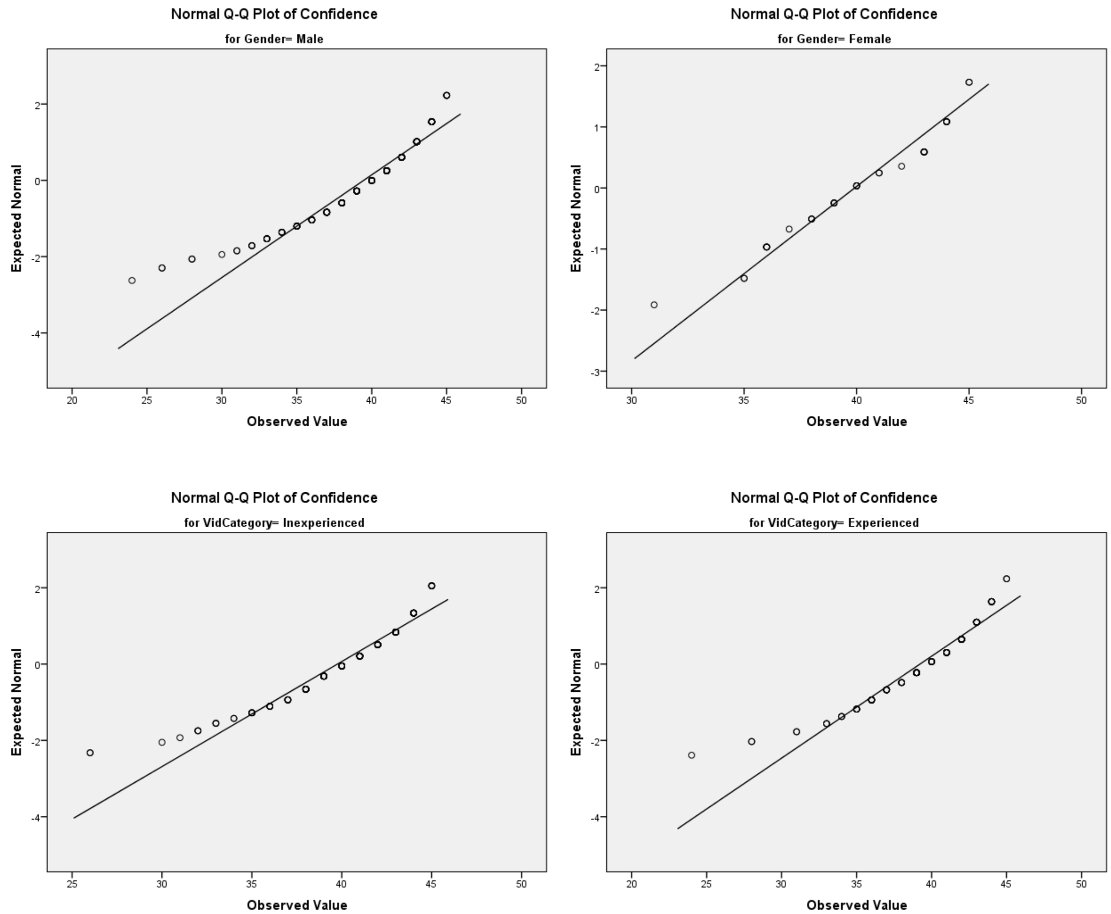


Figure 26. Q-Q plots of confidence by group for visual expeditionary skills training (*N* = 265).

The Q-Q plots for the combined SAR variable are shown in Figure 27. All four showed some kurtosis but the data appeared to be approximately normal. The MANOVA was accomplished but the deviations from normal for confidence were considered with the results.

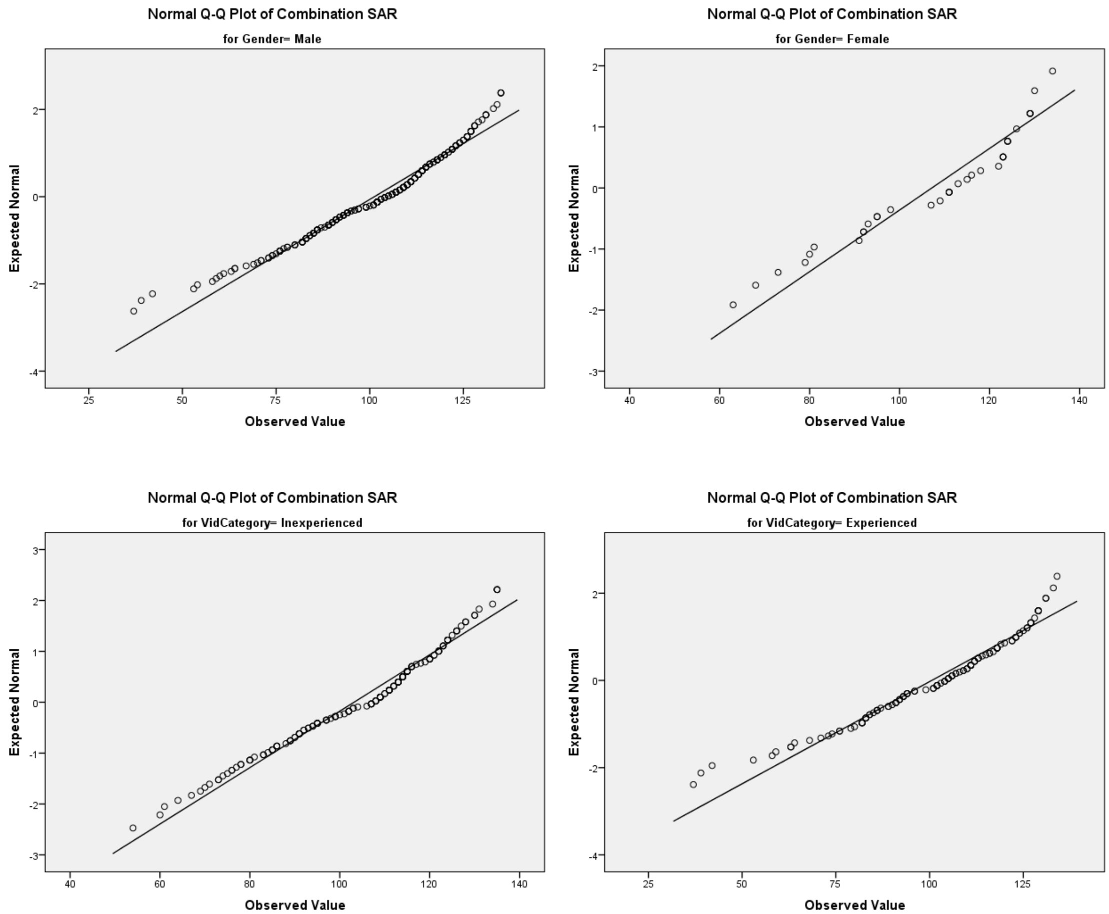


Figure 27. Q-Q plots of satisfaction by group for visual expeditionary skills training (*N* = 265).

Linearity was the fifth assumption and was checked using the scatterplots shown in Figures 28, 29, 30, and 31. The plots were assessed to have a linear relationship between the dependent variables. Of note, the most linearity appeared between perceived effectiveness and the combined SAR variable. The least linearity appeared between usability and perceived effectiveness. This matched the results of the original simulation linearity check.

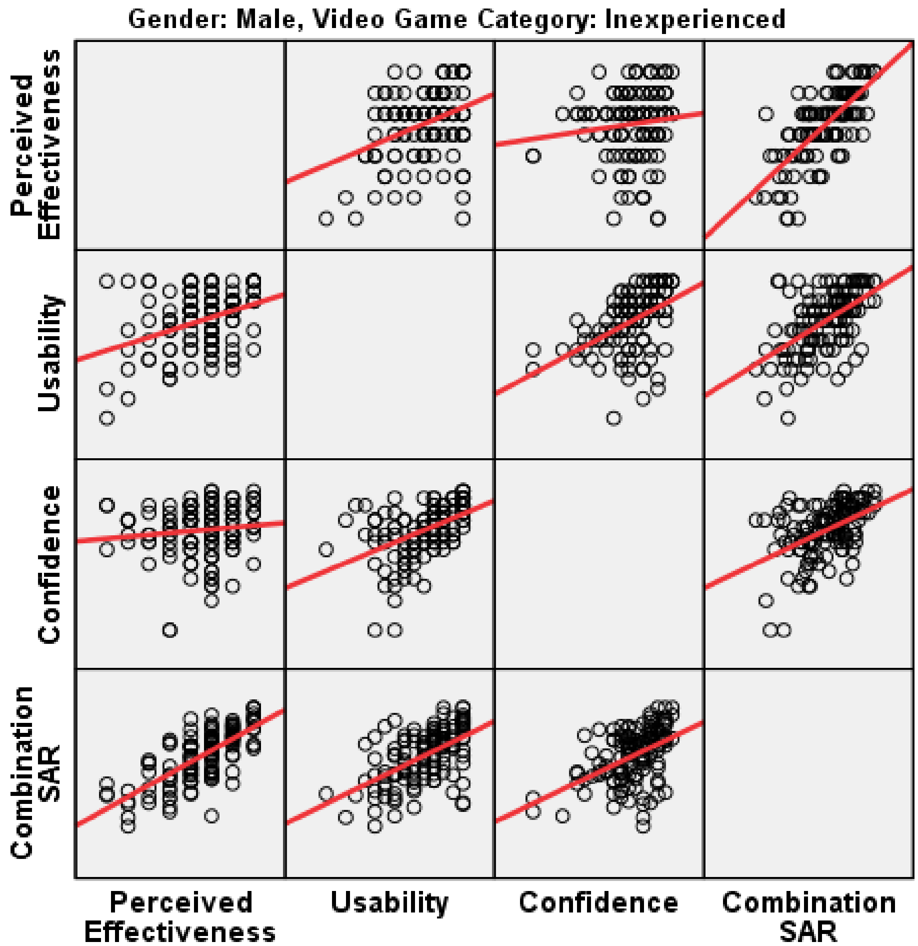


Figure 28. Scatterplot of male inexperienced video gamers for visual expeditionary skills training ($N = 265$).

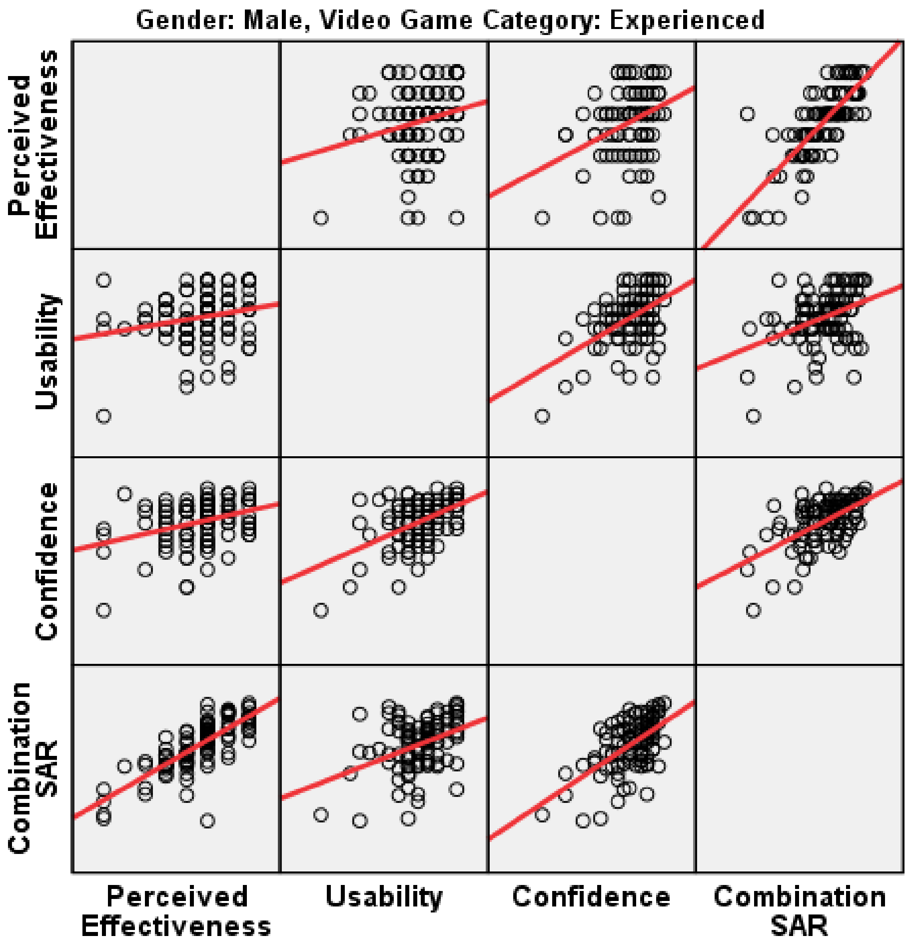


Figure 29. Scatterplot of male experienced video gamers for visual expeditionary skills training ($N = 265$).

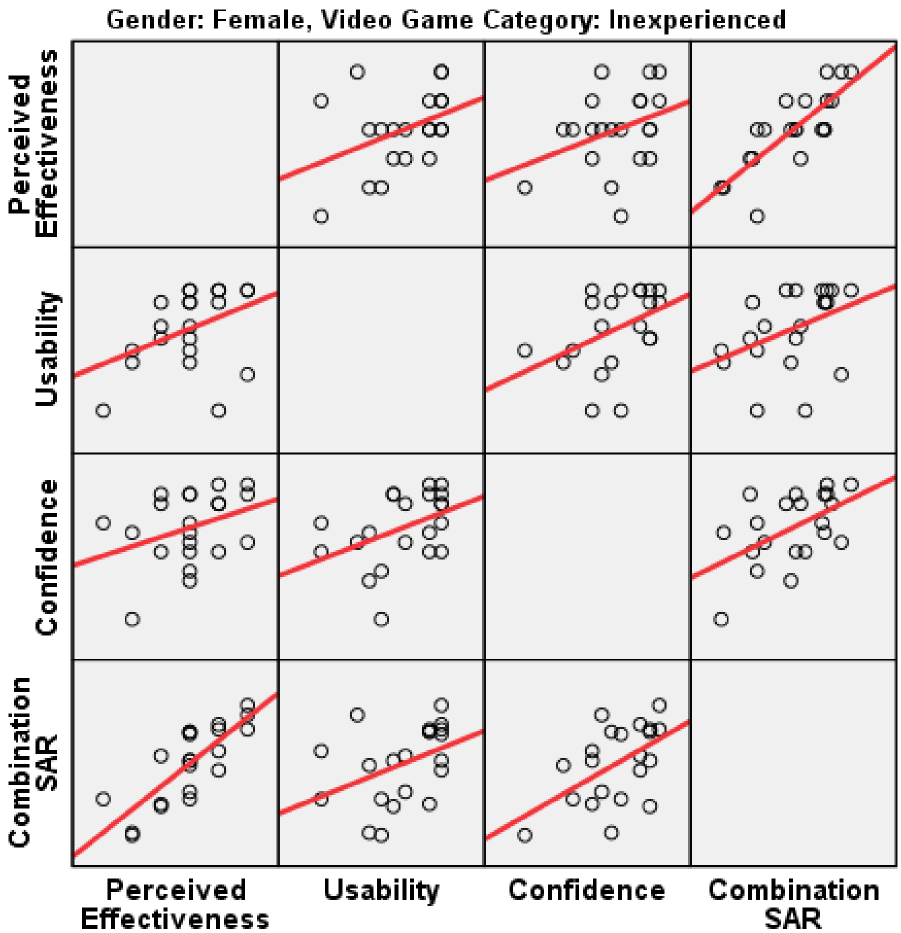


Figure 30. Scatterplot of female inexperienced video gamers for visual expeditionary skills training (N = 265).

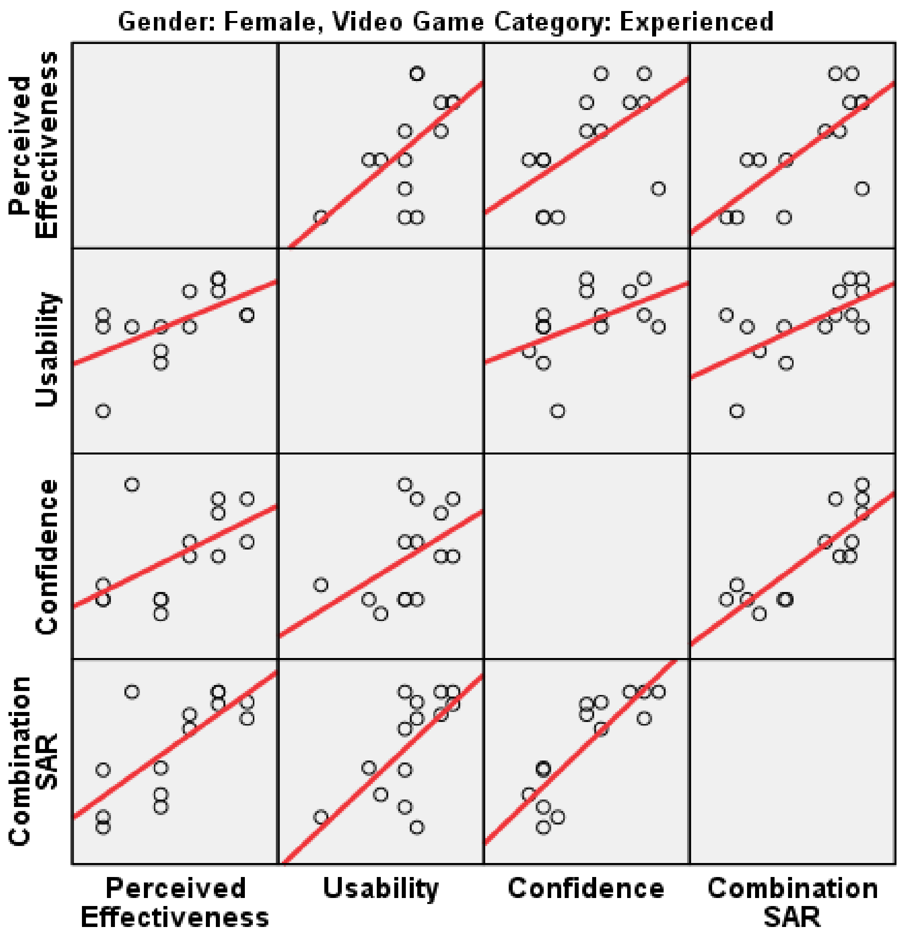


Figure 31. Scatterplot of female experienced video gamers for visual expeditionary skills training ($N = 265$).

The sixth assumption was the existence of the homogeneity of variance-covariance matrices. This was tested using both the Levene's test for individual dependent variables and Box's M for the covariance. Levene's results, as listed in Table 61, were all greater than 0.05, indicating there was homogeneity of variance. Box's M was 0.373, indicating homogeneity of covariance so the sixth assumption was met.

Table 61

Levene's Test of Variances for Visual Expeditionary Skills Training

Variable	Levene Test Original Simulation
Perceived Effectiveness	0.399
Usability	0.136
Confidence	0.929
Satisfaction	0.092

N = 265

The seventh and final assumption was no multicollinearity. This was tested using Pearson's correlation coefficients between each of the dependent variables. The results are listed in Table 62. The lowest Pearson's coefficient was 0.247 between confidence and perceived effectiveness. The highest was 0.7460 between perceived effectiveness and the combined SAR variable. All the coefficients were below 0.90, indicating no multicollinearity in the dataset so the assumption was met.

Table 62

Pearson's Correlation Coefficients Test for Multicollinearity: Visual Expeditionary Skills Training

		Perceived Effectiveness	Usability	Confidence	Combination SAR
Perceived Effectiveness	Pearson Correlation	1	.332**	.247**	.746**
	Sig. (2-tailed)		.000	.000	.000
	N	265	265	265	265
Usability	Pearson Correlation	.332**	1	.481**	.481**
	Sig. (2-tailed)	.000		.000	.000
	N	265	265	265	265
Confidence	Pearson Correlation	.247**	.481**	1	.546**
	Sig. (2-tailed)	.000	.000		.000
	N	265	265	265	265
Combined SAR	Pearson Correlation	.746**	.481**	.546**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	265	265	265	265

** Correlation is significant at the 0.01 level (2-tailed).

The MANOVA assumptions test results shown in Table 63 indicated outlier problems with the database and a small departure from normality for confidence. The outlier problem was dealt with by running a MANOVA test with the outliers present; if there was significance, a second MANOVA would have been conducted with an adjusted dataset with the outliers removed. The small departure from normality was considered in the results.

Table 63

Results of Requirement/Assumption Check for Visual Expeditionary Skills Training

MANOVA Requirements/Assumptions	For Each DV	Within Groups
R1. DV measured as interval	✓	✓
R2. IV are categorical	✓	✓
A1. Independent Observations	✓	✓
A2. Adequate Sample Size	✓	✓
A3. No Outliers	17 Outliers	3 Outliers
A4. Normality	Not for Confidence	Not for Confidence
A5. Linearity	✓	✓
A6. Homogeneity of Variance	✓	✓
A7. Multicollinearity	✓	✓

N = 264

APPENDIX G

MULTIVARIATE ANALYSIS OF VARIANCE ASSUMPTION CHECK FOR AGE: ORIGINAL SIMULATION

This appendix contains a detailed examination of the MANOVA assumptions check for the original simulation dataset using age as the independent variable and perceived effectiveness, usability, confidence, and the combined SAR (satisfaction, attention, and relevance) as the dependent variables.

The sample size was 898 for the original simulation dataset. Appendix E has the details of the initial requirement and assumptions check for the original simulation dataset using gender and video game experience as the independent variables. This appendix only includes details on those assumptions that needed to be rechecked due to changing the independent variable to age.

The two requirements and first assumption--the dependent variable was interval, the independent variable was categorical, and the observations were independent--were all met.

The second assumption was all groups contained an adequate sample size. As shown in Table 64, the smallest group was 13. The minimum size was four based on the four dependent variables so the assumption was met.

Table 64

Sample Size for Original Simulation Groups: Age

Age	Group Size
30-35	13
36-40	380
41-45	314
46-50	136
>50	55

N = 898

The third assumption was there would be no univariate or multivariate outliers (Lund & Lund, 2013). Like many of the remaining assumptions, this assumption needed to be checked both from univariate and multivariate views. To check for univariate outliers, boxplots of each dependent variable were used for each group of independent variables. The results are presented in Figures 32, 33, 34, and 35. Confidence had 20 univariate outliers. The combined SAR variable had one outlier but it was also an outlier for confidence.

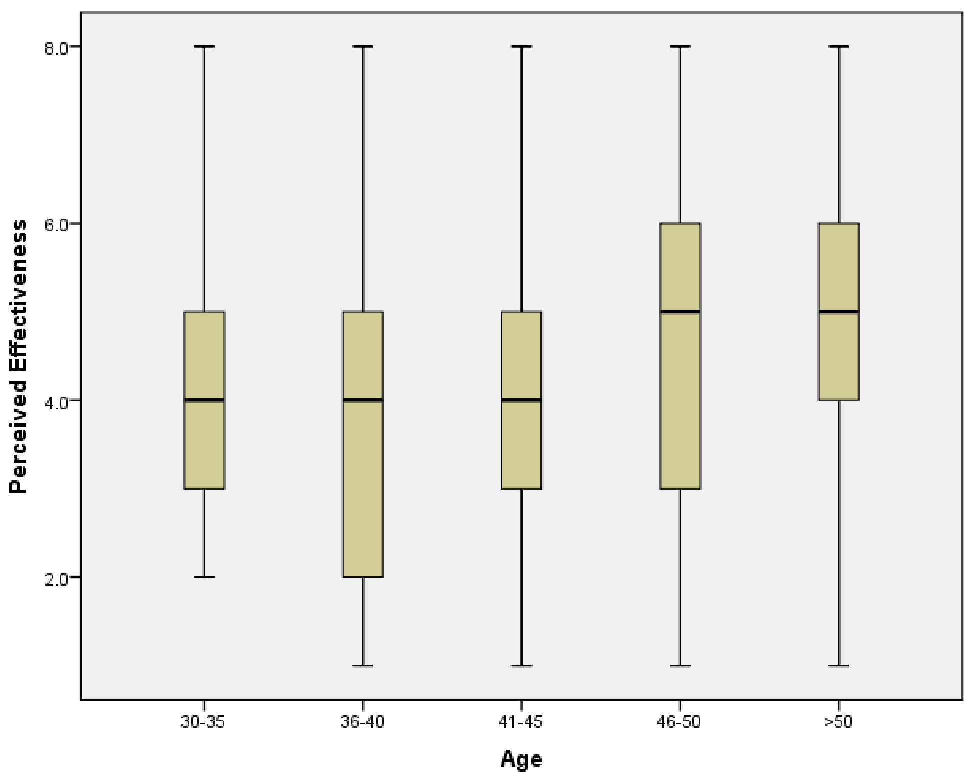


Figure 32. Boxplots of perceived effectiveness: Age (N = 898).

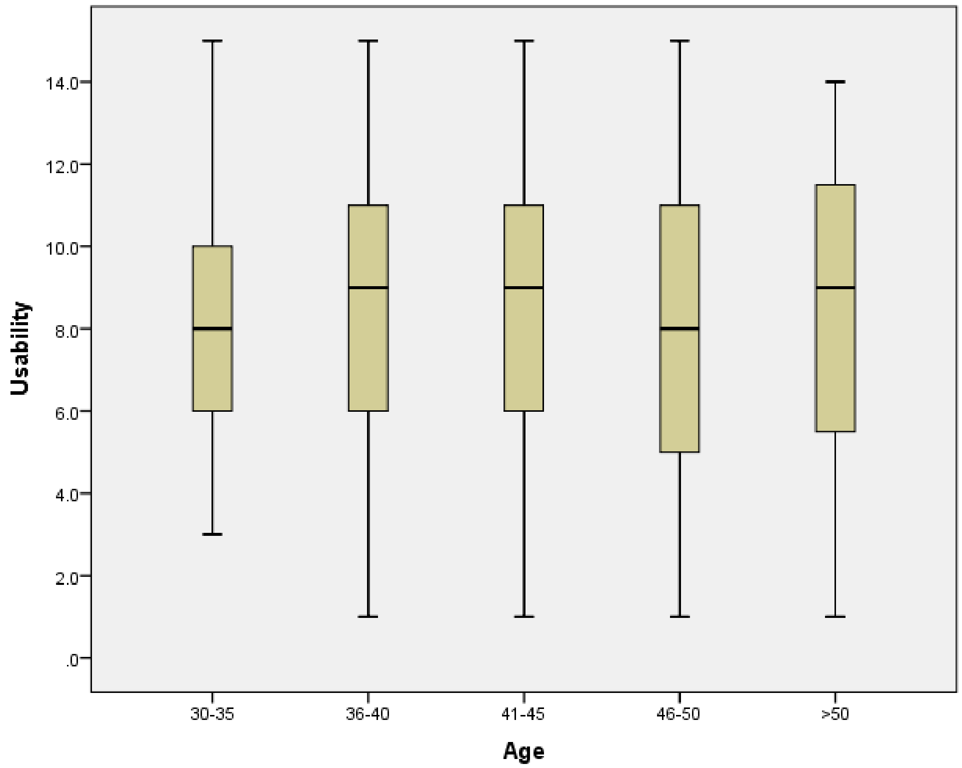


Figure 33. Boxplots of usability: Age ($N = 898$).

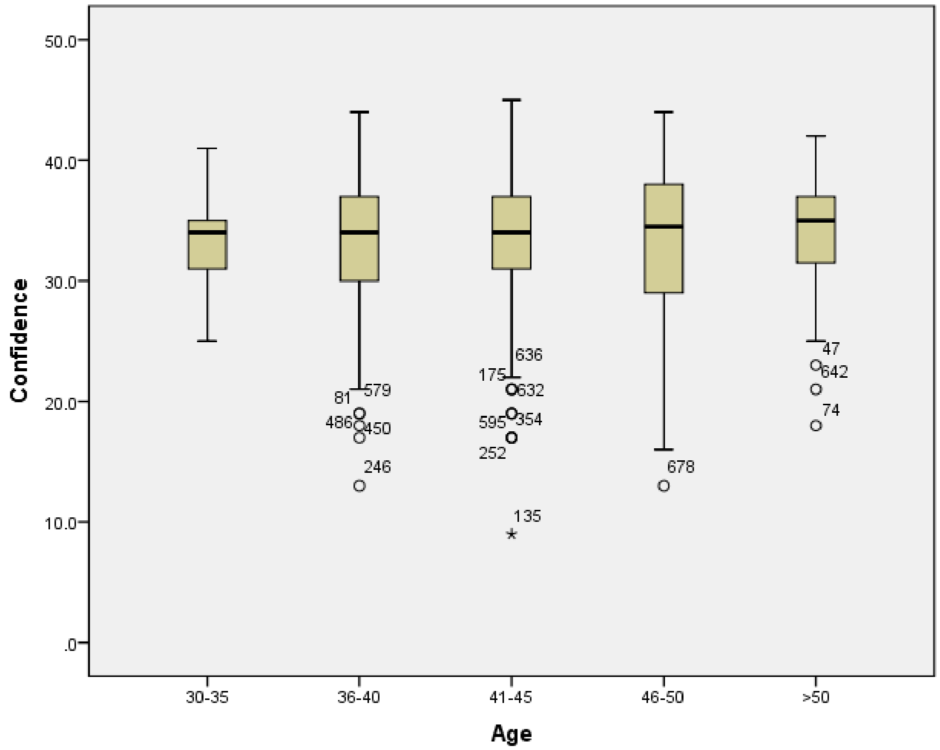


Figure 34. Boxplots of confidence: Age (N = 898).

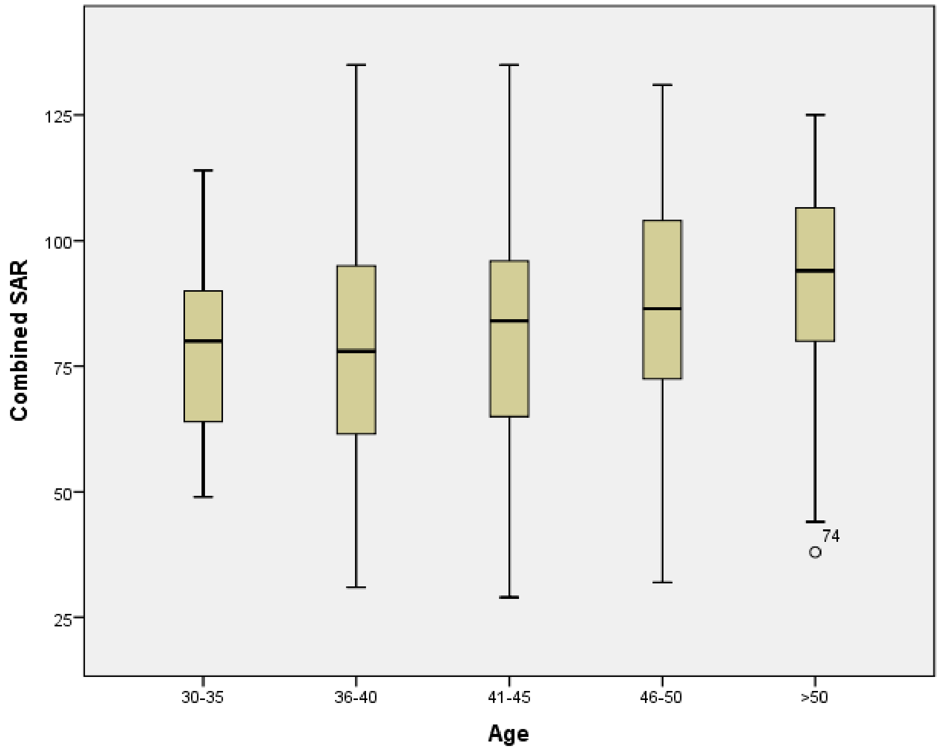


Figure 35. Boxplot of combined satisfaction, attention, and relevance variable: Age (*N* = 898).

There was no indication of data entry error. Seventeen outliers were inexperienced video gamers and three were experienced video gamers. The breakout of outliers is presented in Table 65.

Table 65

Univariate Outliers Broken Out by Group: Age

Age	Number of Outliers
30-35	0
36-40	5
41-45	11
46-50	1
>50	3

To accomplish the check for multivariate outliers, a Mahalanobis distance was created from regression procedures. The Mahalanobis distance was compared against the chi-square distance for the degree of freedom equal to the number of dependent variables-- in this case, four. The critical value of 18.47 was used. Three cases exceeded this value: case numbers 253, 246, and 135. Samples 246 and 135 were univariate outliers but sample 253 was not. The total number of univariate and multivariate outliers was 21.

The fourth assumption was the dependent variables would be normally distributed. The assumption of normality was checked by visually examining the Q-Q plots for each variable by group. Visually inspection of the 20 plots indicated the data appeared approximately normal as shown in Figures 36 through 55.

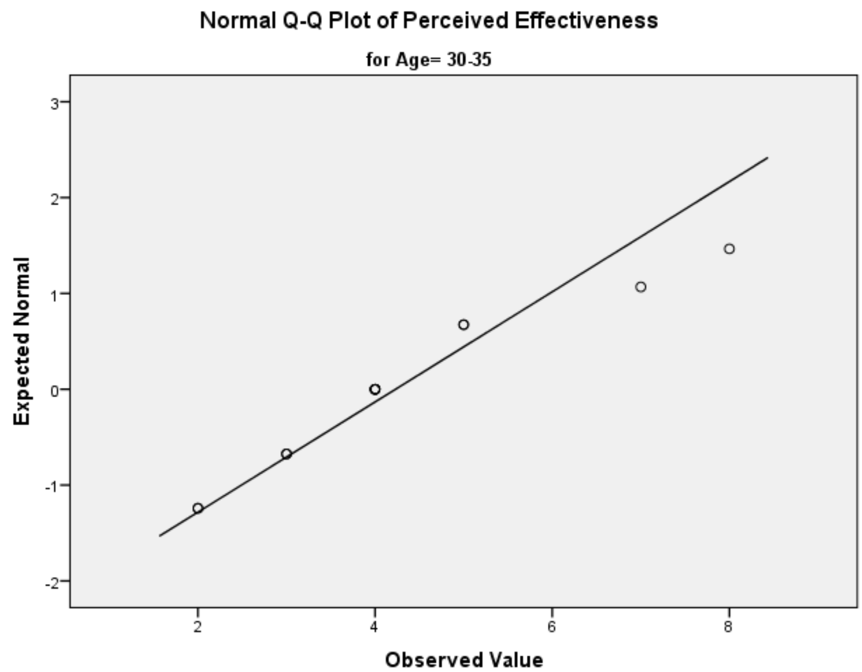


Figure 36. Q-Q plot of perceived effectiveness: Age 30-35 ($N = 898$).

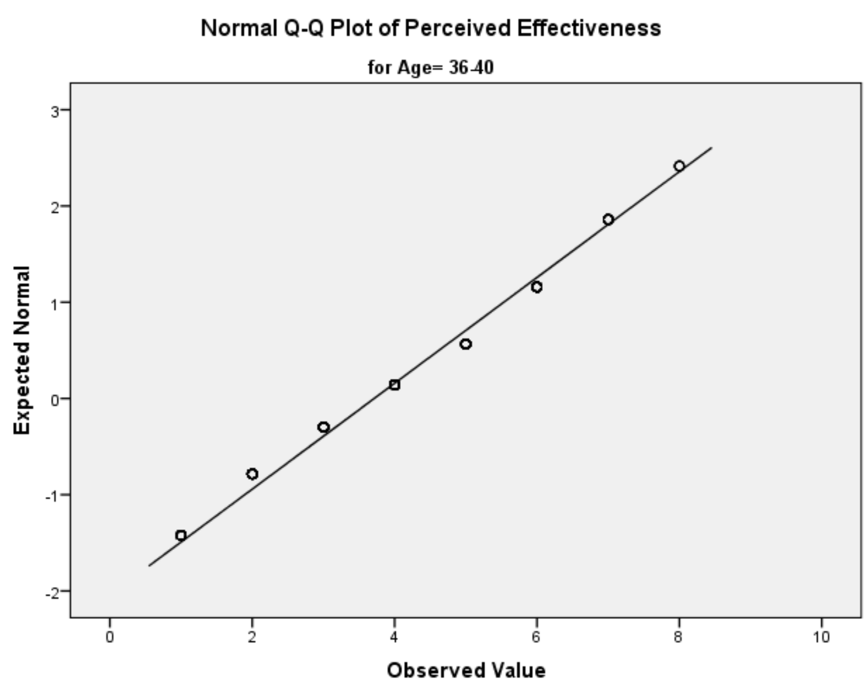


Figure 37. Q-Q plot of perceived effectiveness: Age 36-40 ($N = 898$).

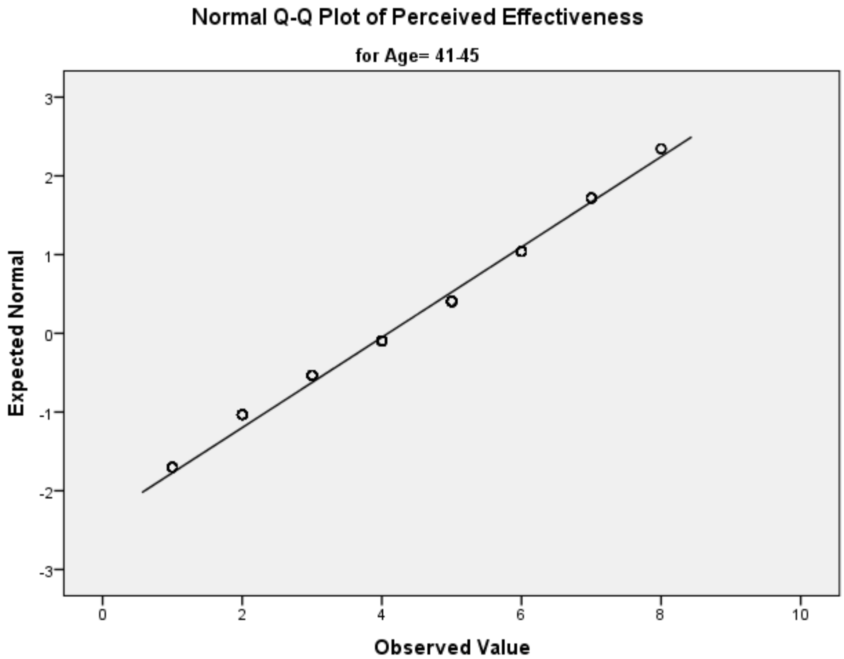


Figure 38. Q-Q plot of perceived effectiveness: Age 41-45 (*N* = 898).

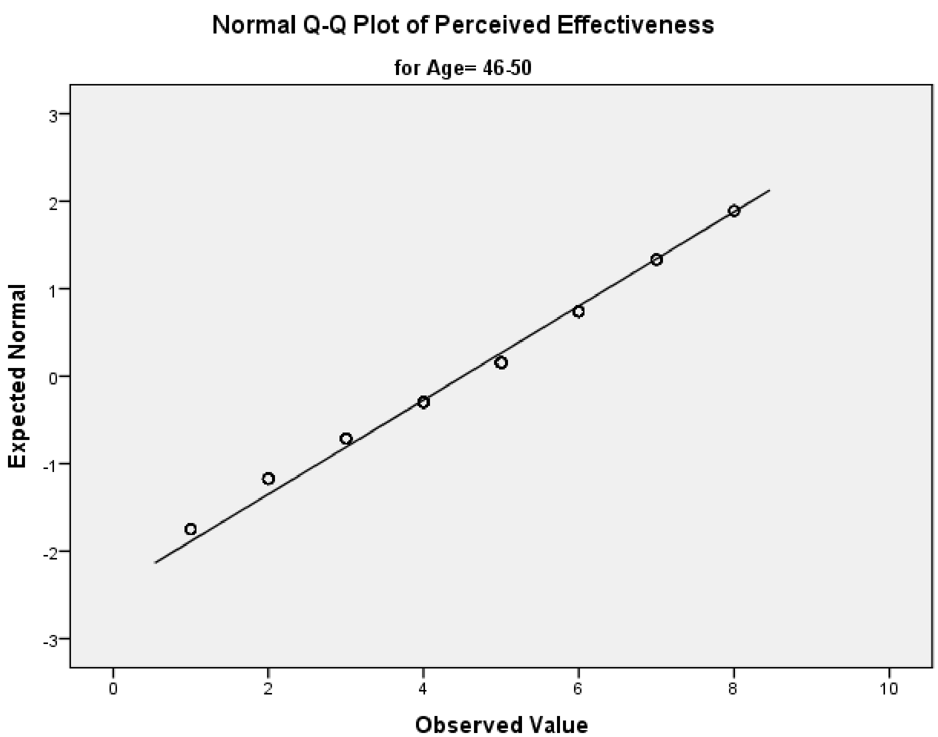


Figure 39. Q-Q plot of perceived effectiveness: Age 46-50 (*N* = 898).

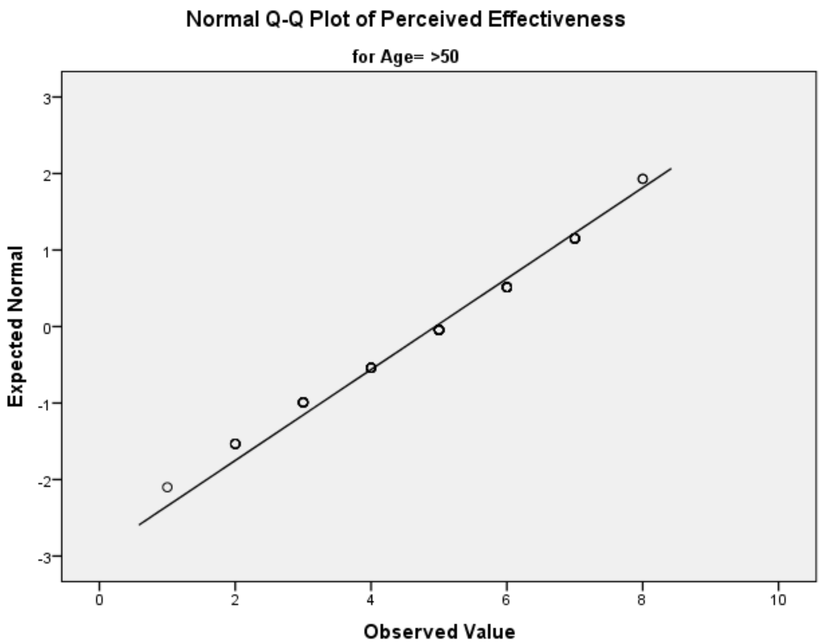


Figure 40. Q-Q plot of perceived effectiveness: Age >50 (N = 898).

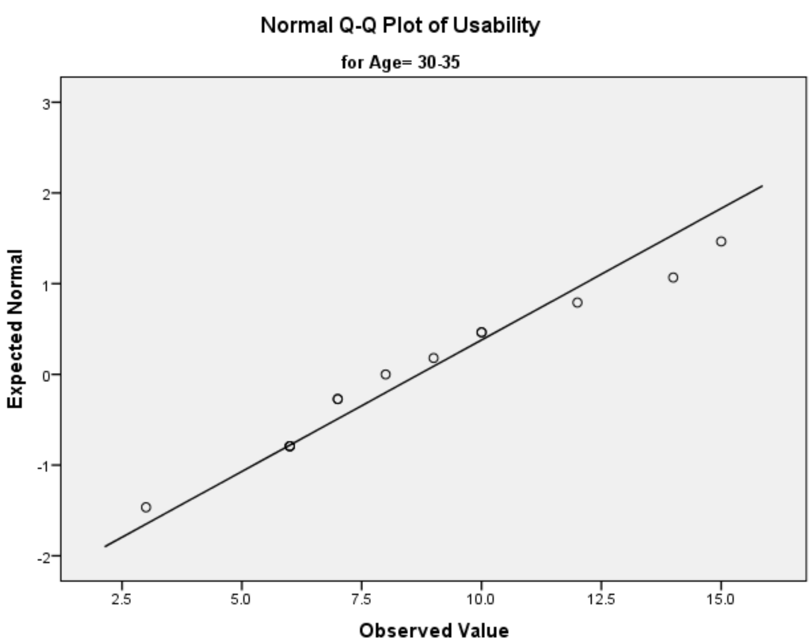


Figure 41. Q-Q plot of usability: Age 30-35 (N = 898).

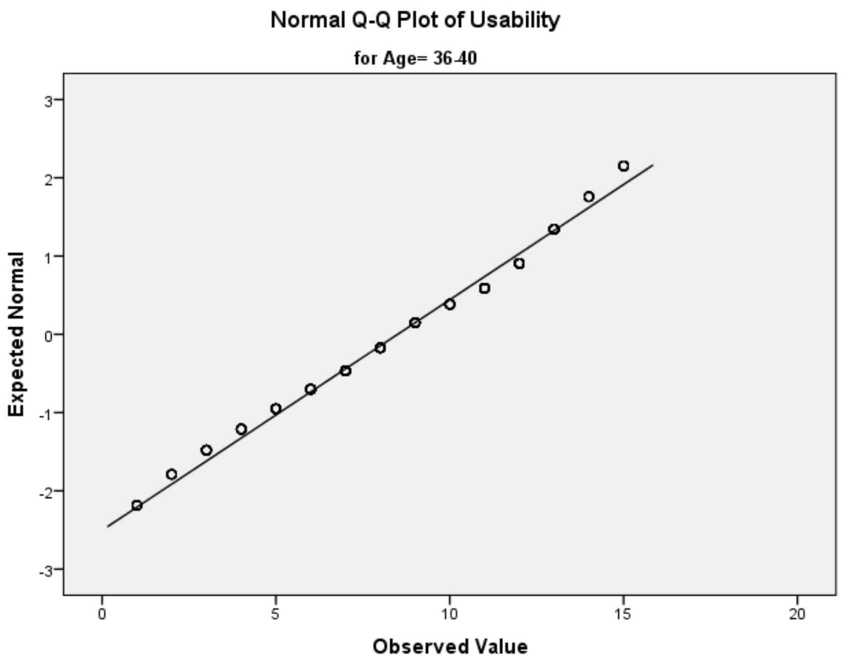


Figure 42. Q-Q plot of usability: Age 36-40 (*N* = 898).

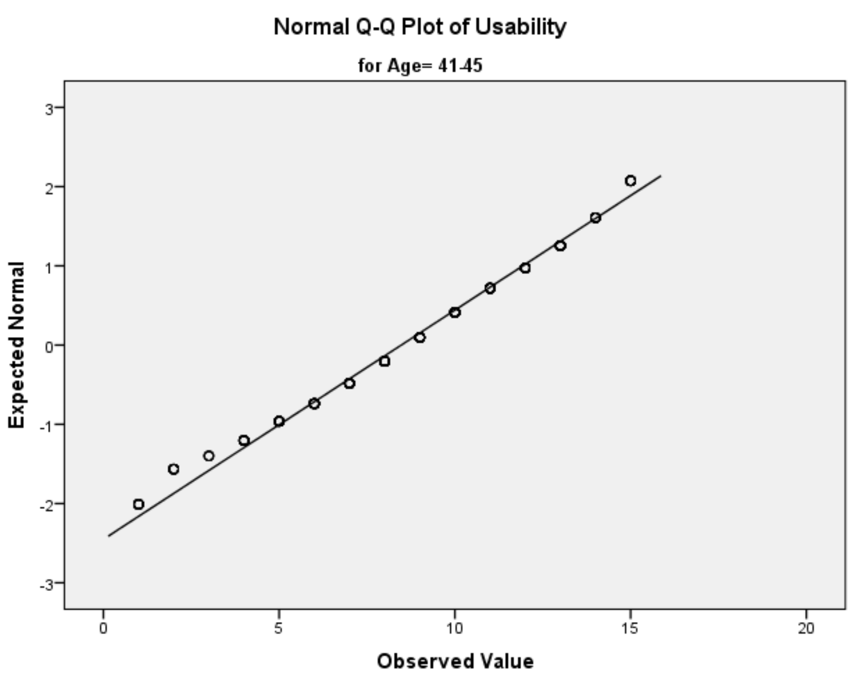


Figure 43. Q-Q plot of usability: Age 41-45 (*N* = 898).

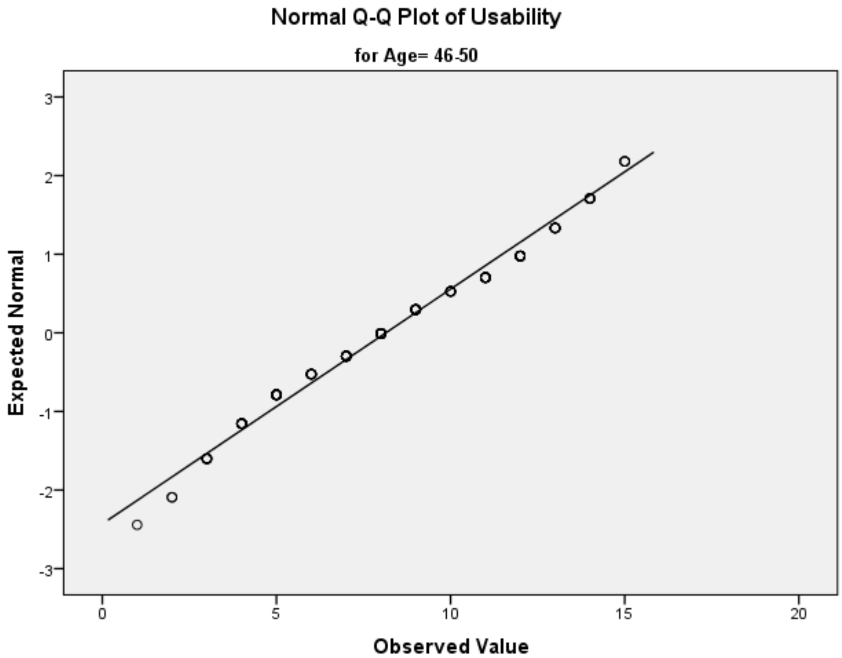


Figure 44. Q Q-Q plot of usability: Age 46-50 (*N* = 898).

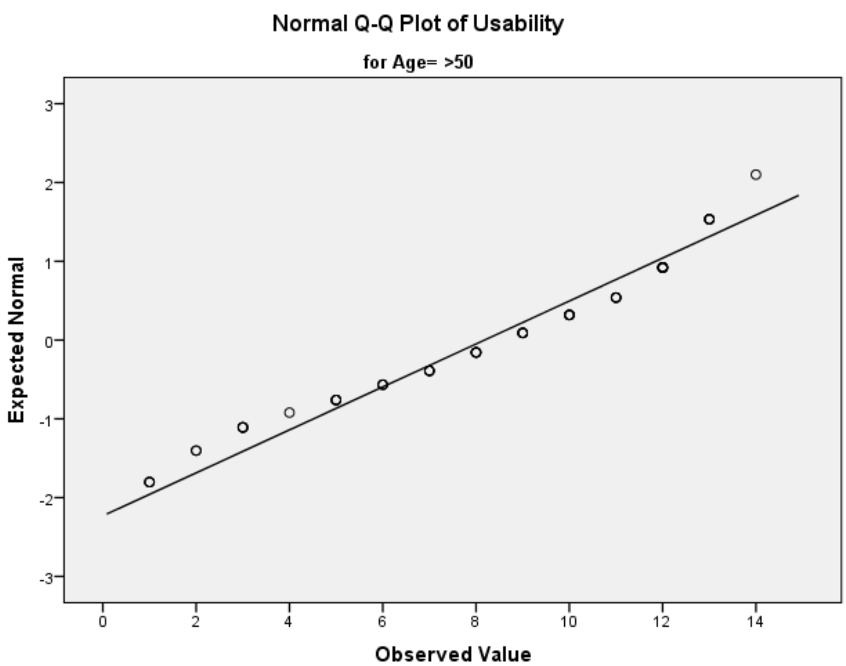


Figure 45. Q-Q plot of usability: Age >50 (*N* = 898).

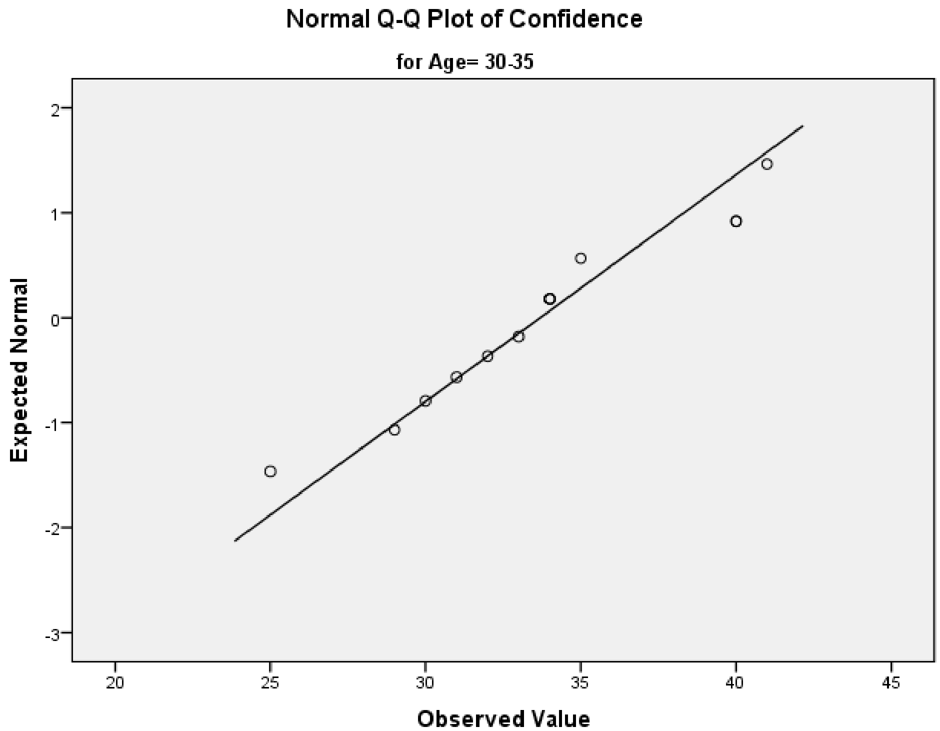


Figure 46. Q-Q plot of confidence: Age 30-35 (*N* = 898).

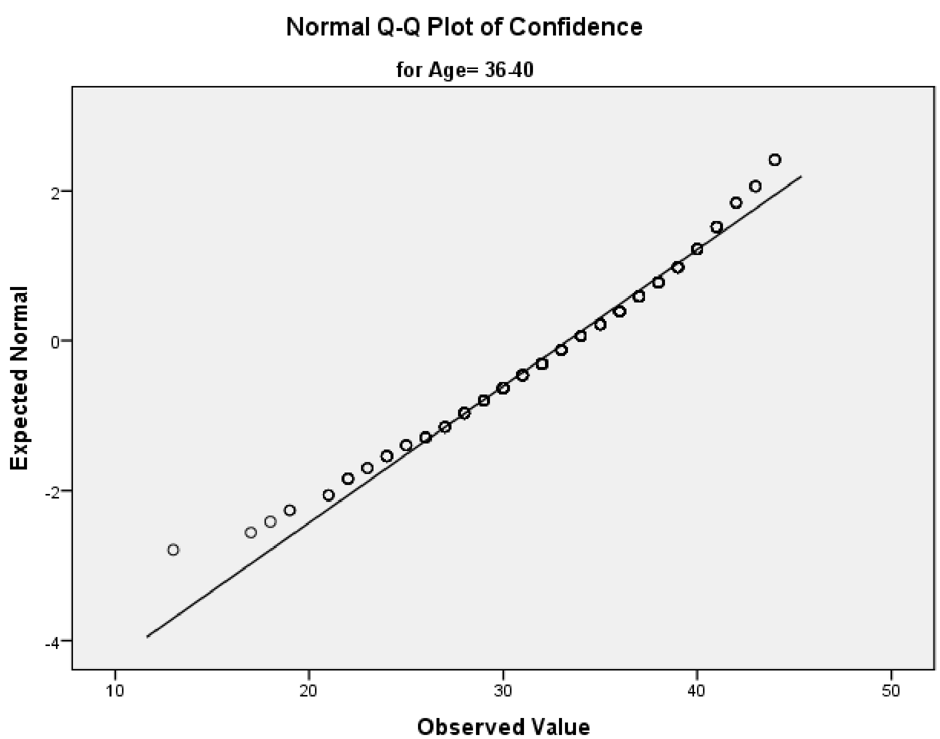


Figure 47. Q-Q plot of confidence: Age 36-40 (*N* = 898).

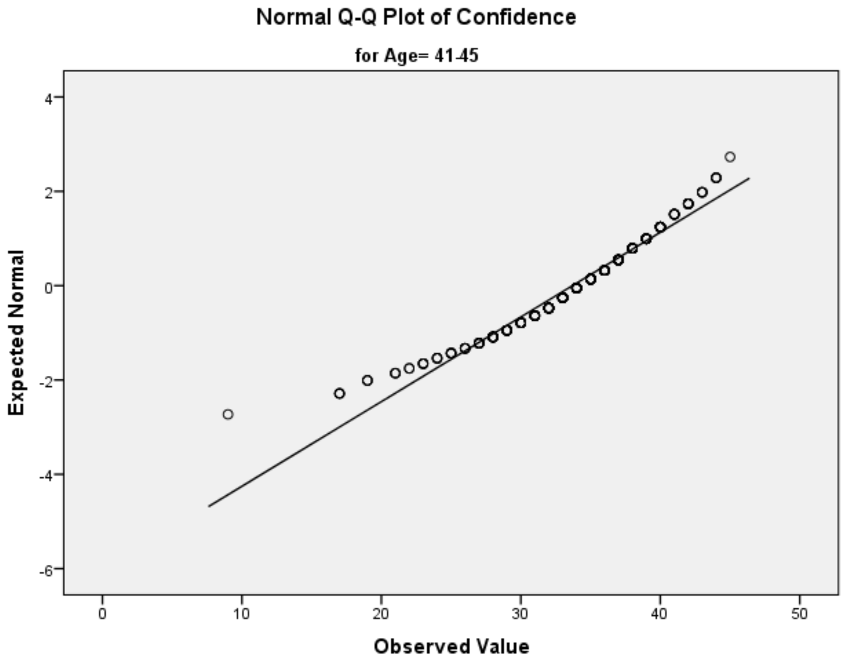


Figure 48. Q-Q plot of confidence: Age 41-45 (N = 898).

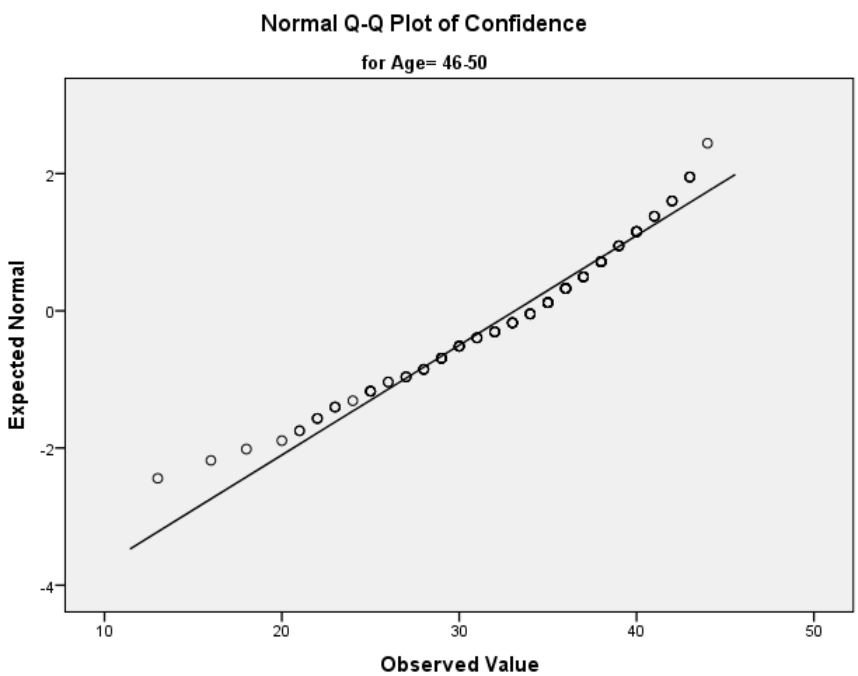


Figure 49. Q-Q plot of confidence: Age 46-50 (N = 898).

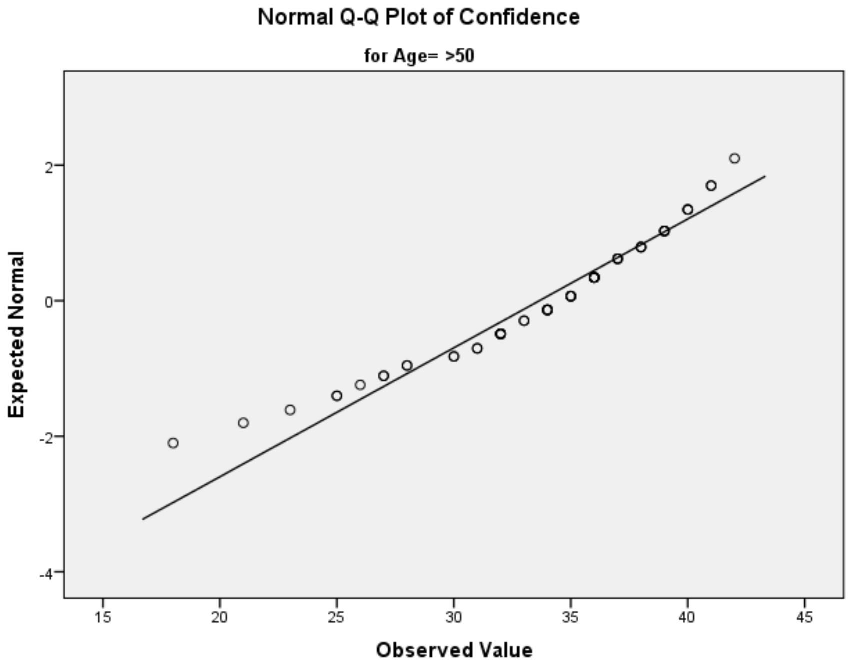


Figure 50. Q-Q plot of confidence: Age > 50 (N = 898).

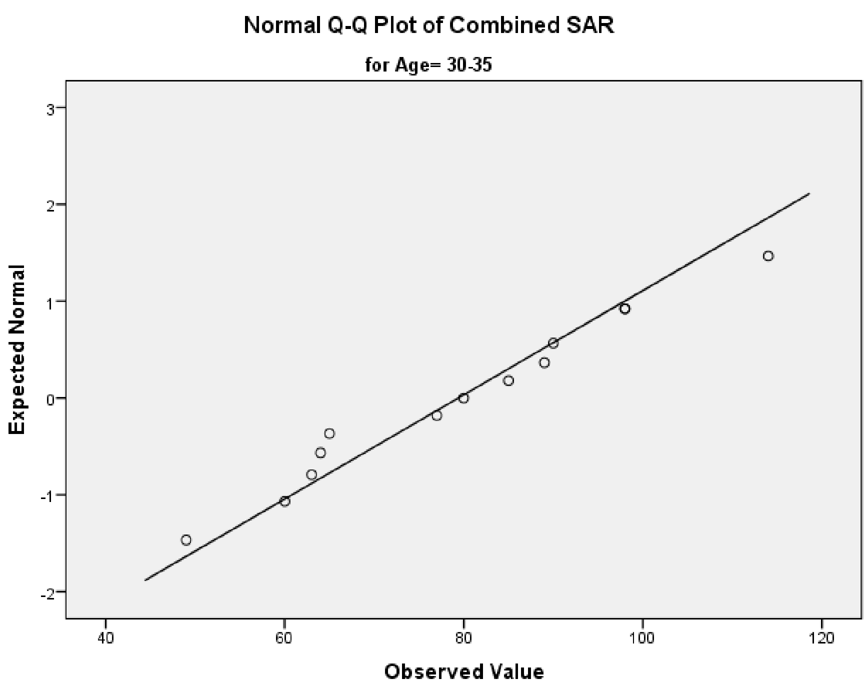


Figure 51. Q-Q plot of combined satisfaction, attention, and relevance: Age 30-35 (N = 898).

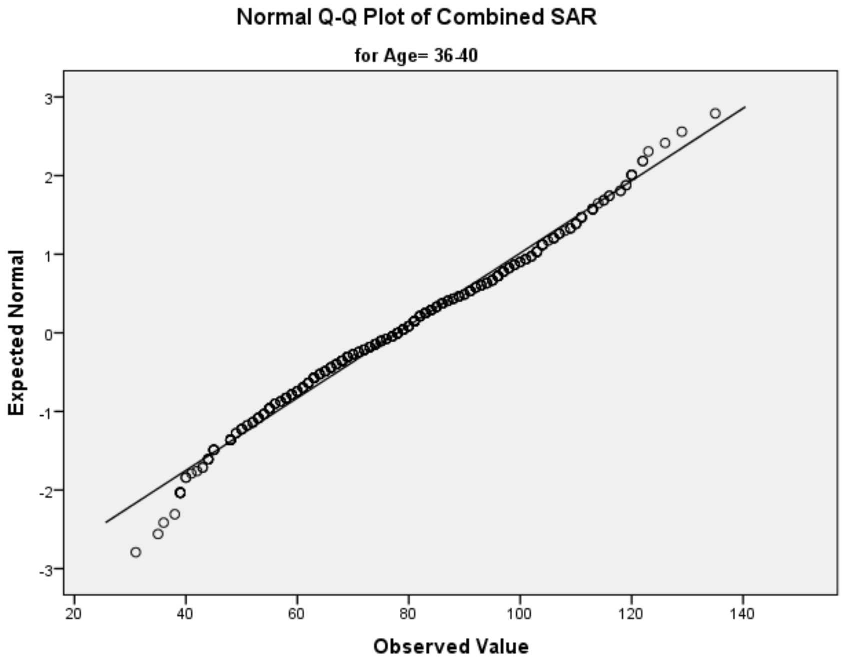


Figure 52. Q-Q plot of combined satisfaction, attention, and relevance: Age 36-40 (N = 898).

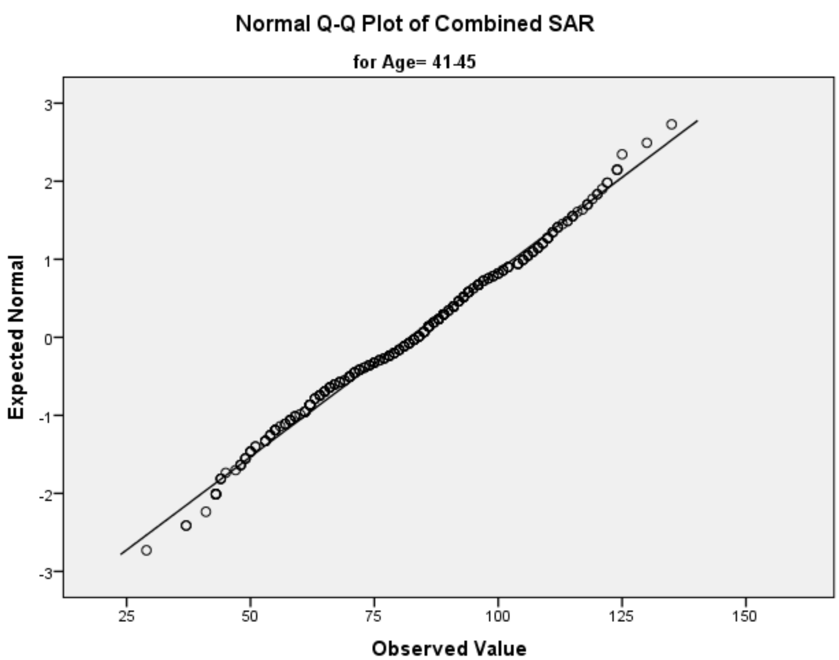


Figure 53. Q-Q plot of combined satisfaction, attention, and relevance: Age 41-45 (N = 898).

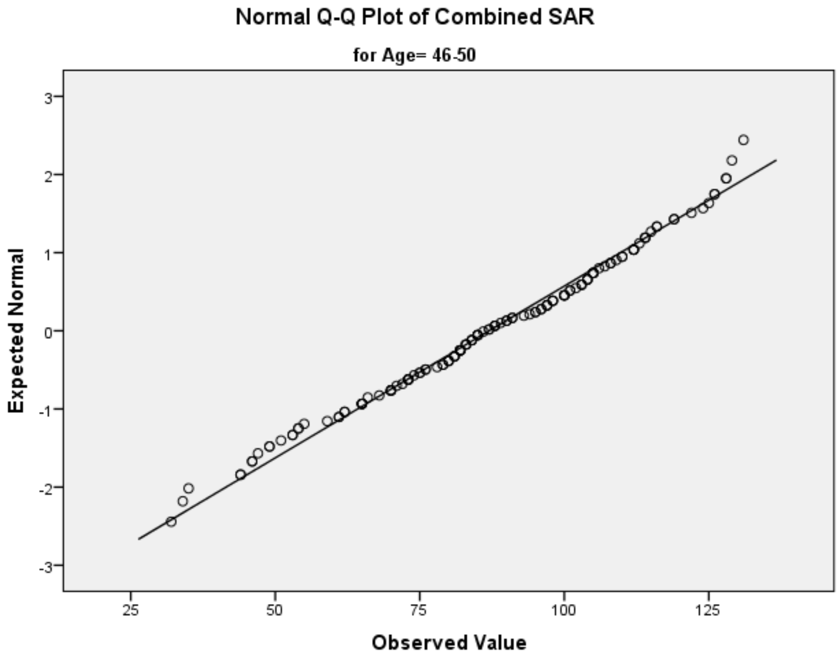


Figure 54. Q-Q plot of combined satisfaction, attention, and relevance: Age 46-50 (*N* = 898).

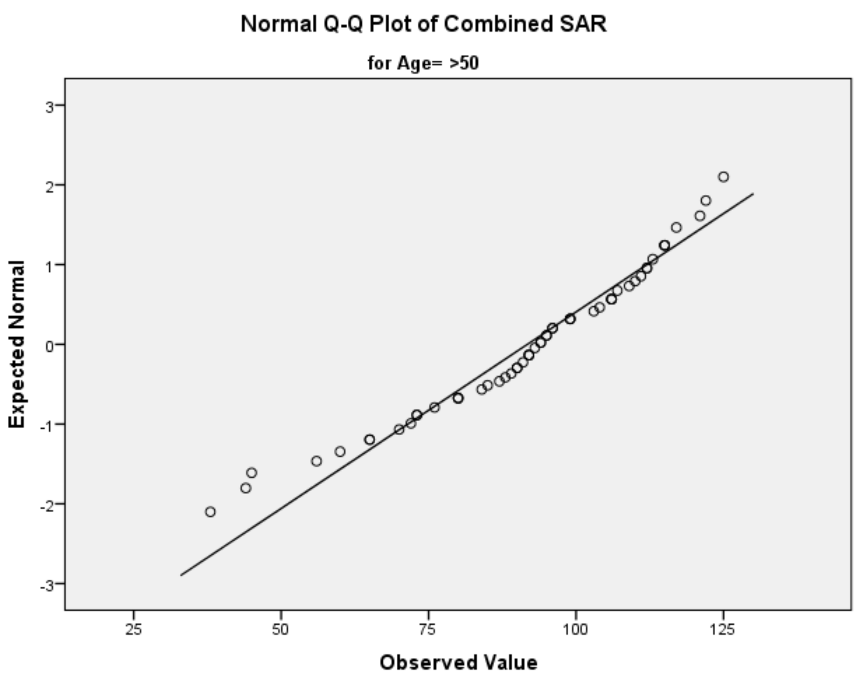


Figure 55. Q-Q plot of combined satisfaction, attention, and relevance: Age > 50 (*N* = 898).

The fifth assumption was linearity, which looked at the relationships between the dependent variables (Lund & Lund, 2013). The scatterplot matrixes shown in Figures 56 through 60 were used to check this assumption. Visually, the scatterplots indicated a linear relationship existed for each of the pairs and so the assumption of linearity was met. A visual inspection indicated the least linearity appeared between perceived effectiveness and usability.

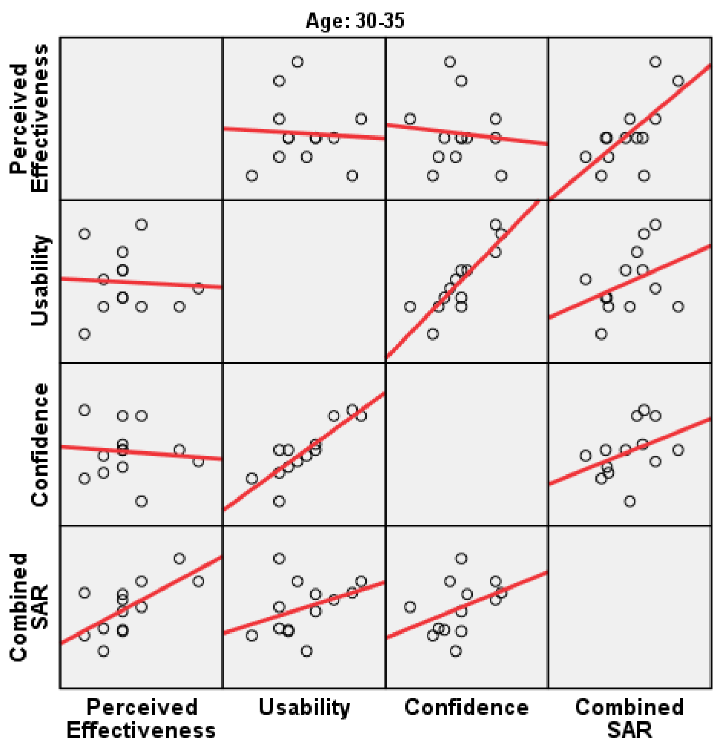


Figure 56. Scatterplot for 30-35 age group ($N = 898$).

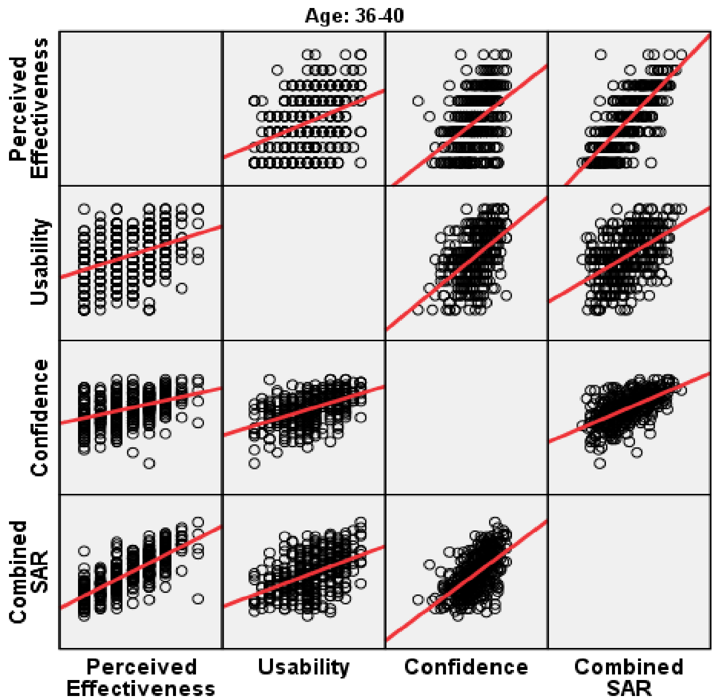


Figure 57. Scatterplot for 36-40 age group ($N = 898$).

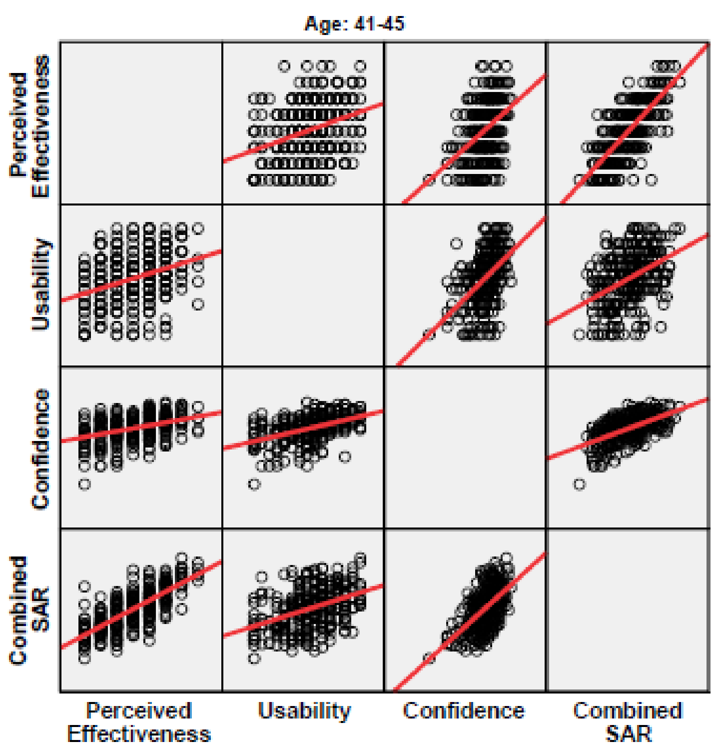


Figure 58. Scatterplot for 41-45 age group ($N = 898$).

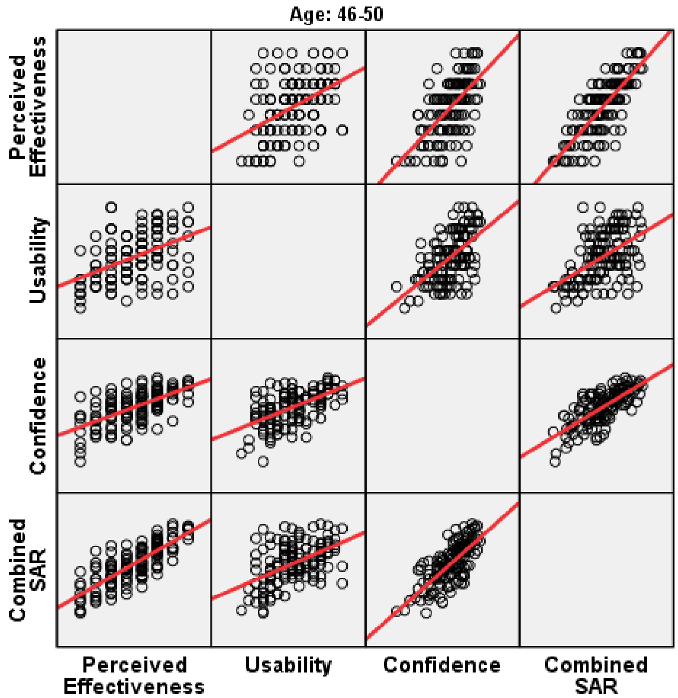


Figure 59. Scatterplot for 46-50 age group (N = 898).

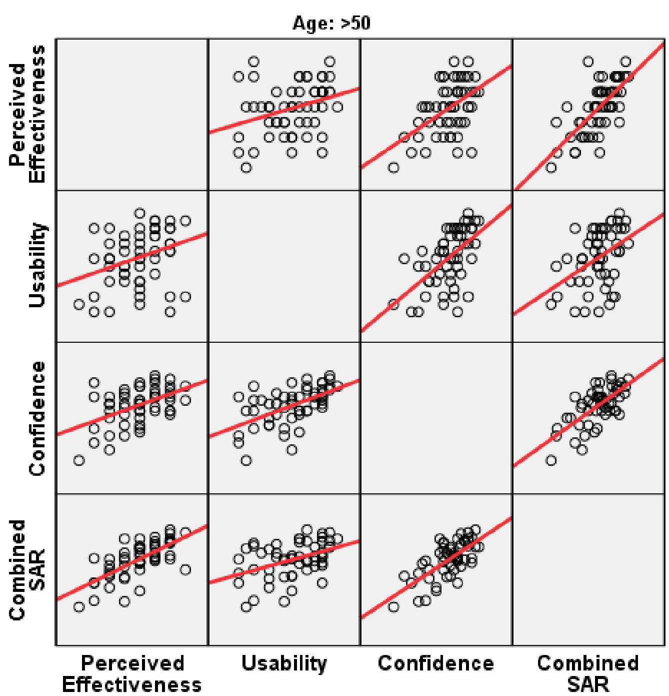


Figure 60. Scatterplot for 50 > age group (N = 898).

The sixth assumption was the existence of the homogeneity of variance and covariances. This was tested using both the Levene's test for individual dependent variables and Box's M for the covariance. Both tests are designed so if significance is found (in this case, less than 0.05), it indicates the sample does not meet the assumption and the variances need to be treated as unequal. Values of greater than .05 mean the variables met the assumption. The values from the Levene's test for each dependent variable are listed in Table 66. All four variables met the requirements for assuming homogeneity of variances. The check for homogeneity of variances within groups (covariance) was checked using Box's M, which had a *p* value of 0.662, so the null hypothesis was rejected and homogeneity of covariances existed across the groups.

Table 66

Levene's Test of Variances for Age

Variable	Levene's Test Original Simulation
Perceived Effectiveness	0.303
Usability	0.906
Confidence	0.139
SAR combined	0.458

N = 898

The seventh and final assumption was no multicollinearity. The dependent variables should be moderately correlated. If they are too low, there is no reason to do the MANOVA and if they are too high (> 0.90), it would be indicate multicollinearity, which would be problematic for running a MANOVA (Lund & Lund, 2013). Multicollinearity was checked using Pearson's correlation coefficients between each of the dependent variables. The results are listed in Table 67. All the coefficients indicated moderate correlations with a low of 0.353 for perceived effectiveness and usability and a high of 0.774 for perceived effectiveness and the combined SAR variable. All pairs showed statistical significance using an alpha of 0.01. Overall, the data met the assumption of no mulitcollinearity.

Table 67

Multicollinearity Check Using Pearson's Coefficients for Age

		Perceived Effectiveness	Usability	Confidence	Combined SAR
Perceived Effectiveness	Pearson Correlation	1	.353**	.445**	.774**
	Sig. (2-tailed)		.000	.000	.000
	N	898	898	898	898
Usability	Pearson Correlation	.353**	1	.523**	.448**
	Sig. (2-tailed)	.000		.000	.000
	N	898	898	898	898
Confidence	Pearson Correlation	.445**	.523**	1	.605**
	Sig. (2-tailed)	.000	.000		.000
	N	898	898	898	898
Combined SAR	Pearson Correlation	.774**	.448**	.605**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	898	898	898	898

** Statistical significance using an alpha of 0.01. $N = 898$

Except for outliers, all MANOVA assumptions were met with this dataset as shown in Table 68. Twenty univariate outliers existed for the confidence variable. The combined SAR variable had one univariate outlier but it was also an outlier for the confidence variable. There were three multivariate outliers but two of those were also in the group of univariate outliers. Therefore, there were 21 outliers total. The outliers seemed to be valid numbers and 21 outliers of 898 samples was only 2.3% of the surveys. However, a MANOVA was run with and without the outliers to test the effect they had on the results.

Table 68

Results of Assumption Check

MANOVA Assumptions	Univariate	Multivariate
1. DV Measured as Interval	✓	✓
2. IV Are Categorical	✓	✓
3. Independent Observations	✓	✓
4. Adequate Sample Size	✓	✓
5. No Outliers	20 identified	3 identified
6. Normality	✓	✓
4. Linearity	✓	✓
5. Homogeneity of Variance	✓	✓
6. Multicollinearity	✓	✓

APPENDIX H

MULTIVARIATE ANALYSIS OF VARIANCE ASSUMPTION CHECK FOR AGE: VISUAL EXPEDITIONARY SKILLS TRAINING

This appendix contains a detailed examination of the MANOVA assumptions check for the VEST dataset using age as the independent variable and perceived effectiveness, usability, confidence, and the combined SAR (satisfaction, attention, and relevance) as dependent variables.

The sample size was 265 for the VEST dataset. Appendix E has the details of the initial requirement and assumptions check using gender and video game experience as the independent variables. This appendix only includes details on those assumptions that needed to be rechecked due to changing the independent variable to age.

The two requirements and first assumption--that the dependent variable was interval, the independent variable was categorical, and the observations were independent--were all met.

The second assumption was the groups needed to have an adequate sample size. As shown in Table 69, the smallest group was seven, which was larger than the number of dependent variables, so the assumption was met.

Table 69

Sample Size for Visual Expeditionary Skills Training Groups

Age	Group Size
30-35	7
36-40	106
41-45	91
46-50	42
>50	19

The third assumption was there would be no univariate or multivariate outliers (Lund & Lund, 2013). To check for univariate outliers, boxplots of each dependent variable were used for each group of independent variables. The results are presented in Figures 61 through 64, showing outliers are present for each of the four variables.

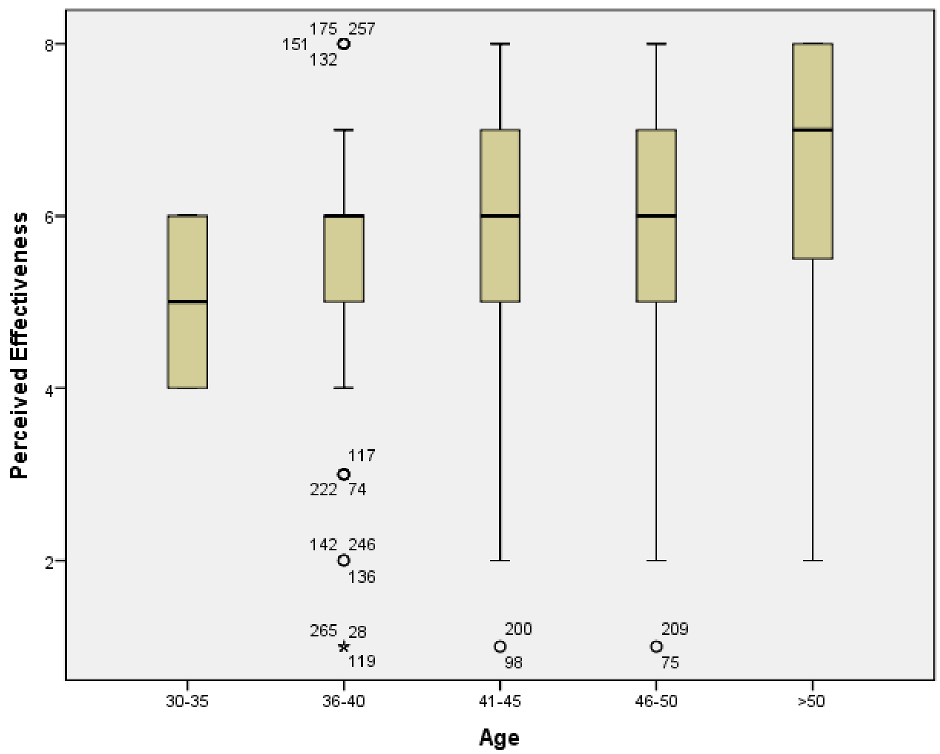


Figure 61. Boxplot of perceived effectiveness by age group (N = 265).

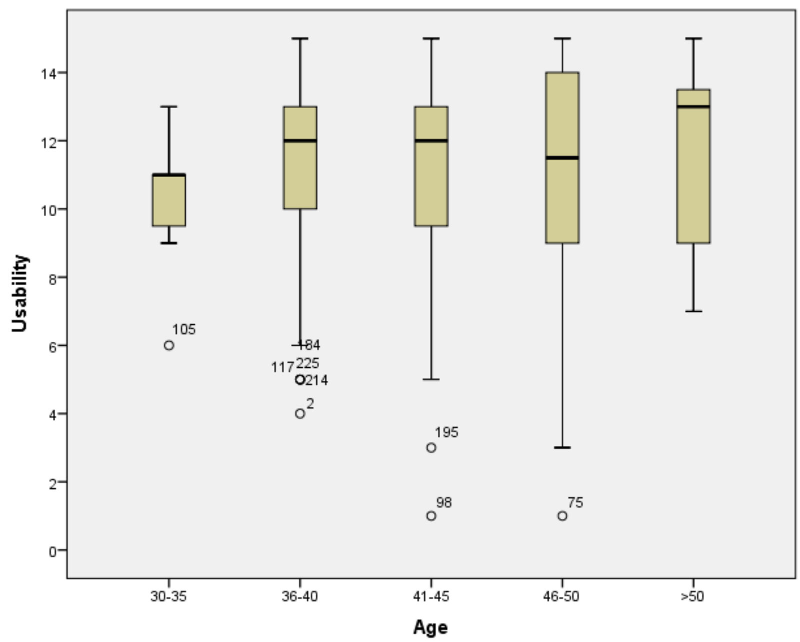


Figure 62. Boxplot of usability by age group ($N = 265$).

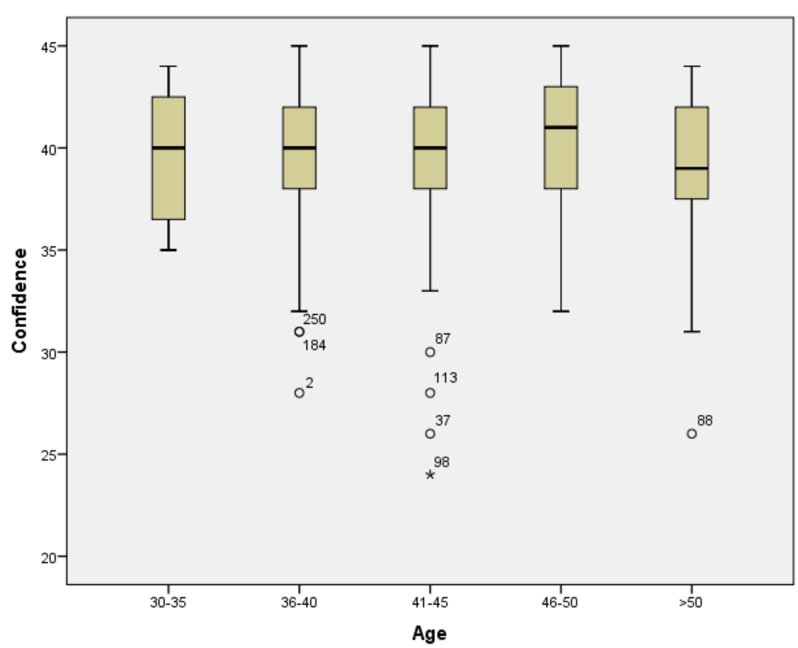


Figure 63. Boxplot of confidence by age group ($N = 265$).

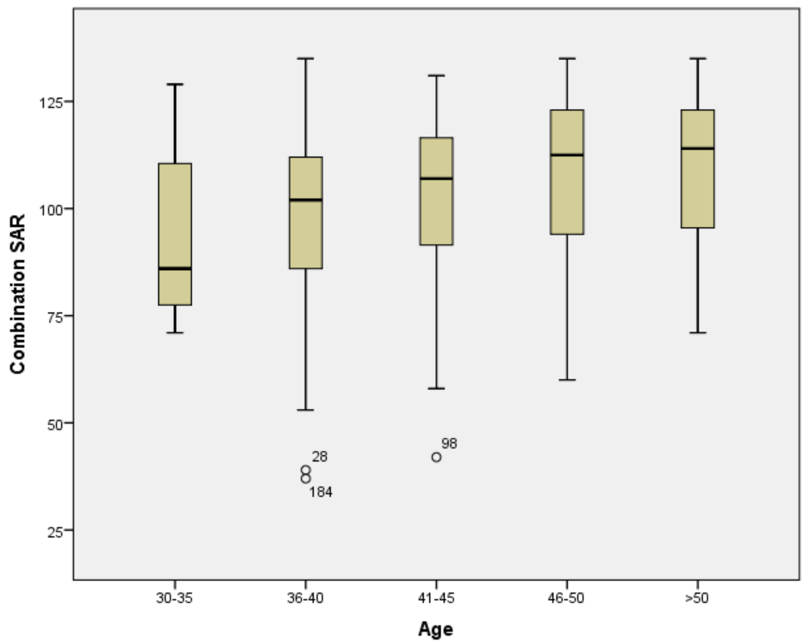


Figure 64. Boxplot of combined satisfaction, attention, and relevance variable by age group (*N* = 265).

There were 49 outlier values across the four variables. However, some samples had outliers for multiple variables. Forty samples contained outliers. Of note, there were nine high value outliers for perceived effectiveness. All the previous outliers for the various MANOVAs had been low value outliers. There was no indication of data entry errors. Nearly all of the outlier values would not have been outliers for the original simulation dataset. However, the smaller VEST dataset and smaller standard deviations resulted in these values being outliers. A breakout of outliers by variable is presented in Table 70.

Table 70

Univariate Outliers Broken Out by Group

Age	Perceived Effectiveness	Usability	Confidence	Combined SAR	Total Outliers	Total Unique Sample Outliers
30-35	0	1	0	0	1	1
36-40	25	5	3	2	35	30
41-45	2	2	4	1	9	6
46-50	2	1	0	0	3	2
>50	0	0	1	0	1	1
				Total	49	40

To accomplish the check for multivariate outliers, a Mahalanobis distance was created from regression procedures. A critical value of 18.47 was used. Three cases exceeded this value: case numbers 184, 98, and 75. All three samples also contained univariate outliers. Forty samples contained univariate or mulitvariate outliers.

The fourth assumption was the dependent variables would be normal. The assumption of normality was checked by visually examining the Q-Q plots for each variable by group. The plots are shown in Figures 65 through 84. The visual appearance of the curves indicated while there were small departures from normal, the data could be considered nearly normal.

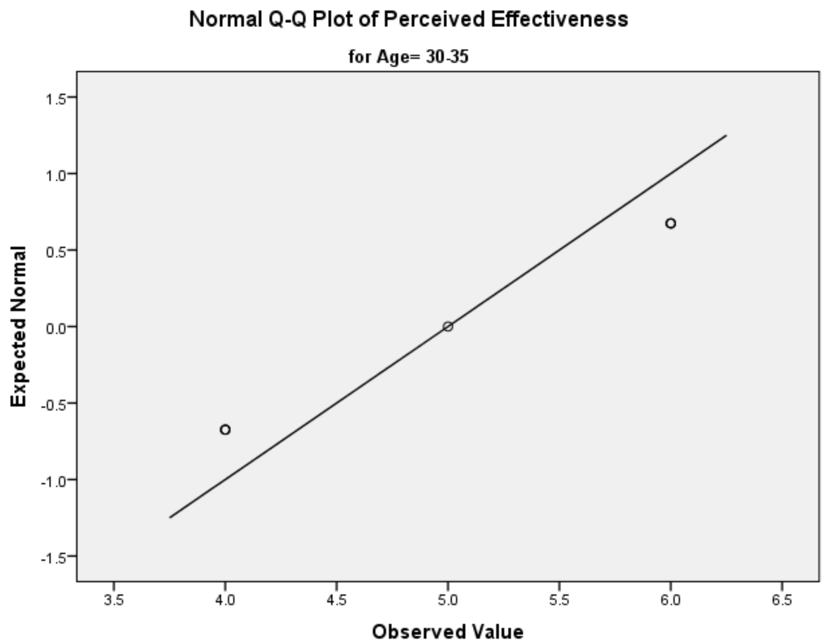


Figure 65. Q-Q plot of perceived effectiveness for age 30-35: Visual expeditionary skills training.

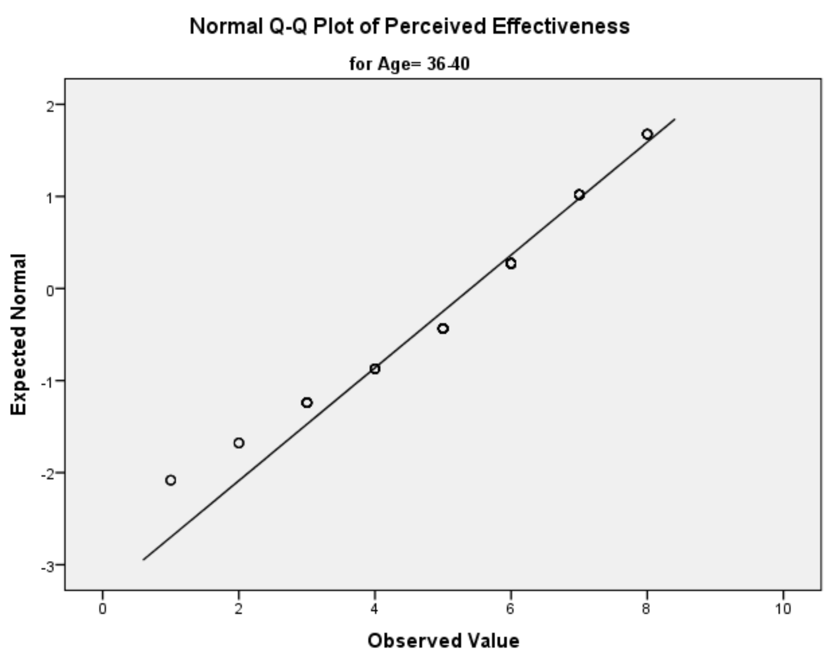


Figure 66. Q-Q plot of perceived effectiveness for age 36-40: Visual expeditionary skills training.

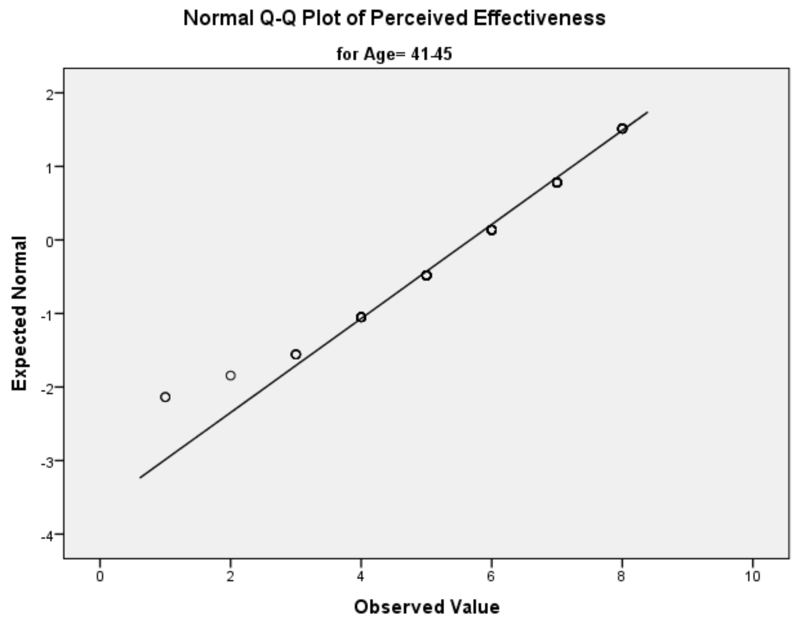


Figure 67. Q-Q plot of perceived effectiveness for age 40-45: Visual expeditionary skills training.

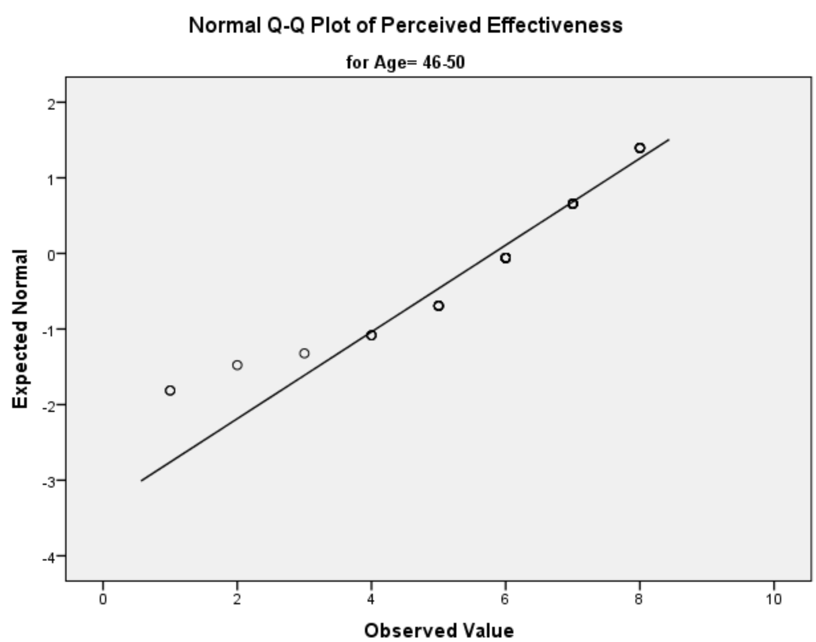


Figure 68. Q-Q plot of perceived effectiveness for age 46-50: Visual expeditionary skills training.

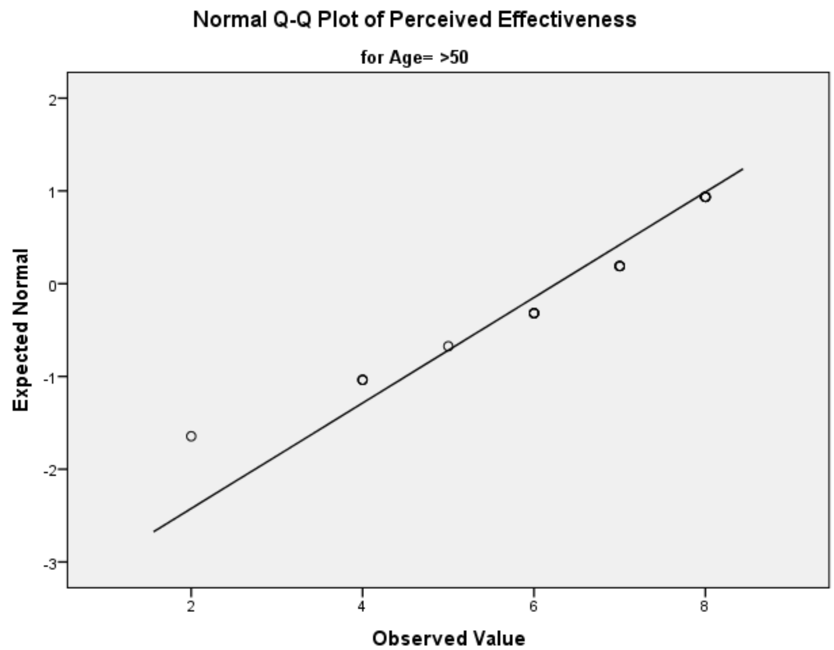


Figure 69. Q-Q plot of perceived effectiveness for age >50: Visual expeditionary skills training.

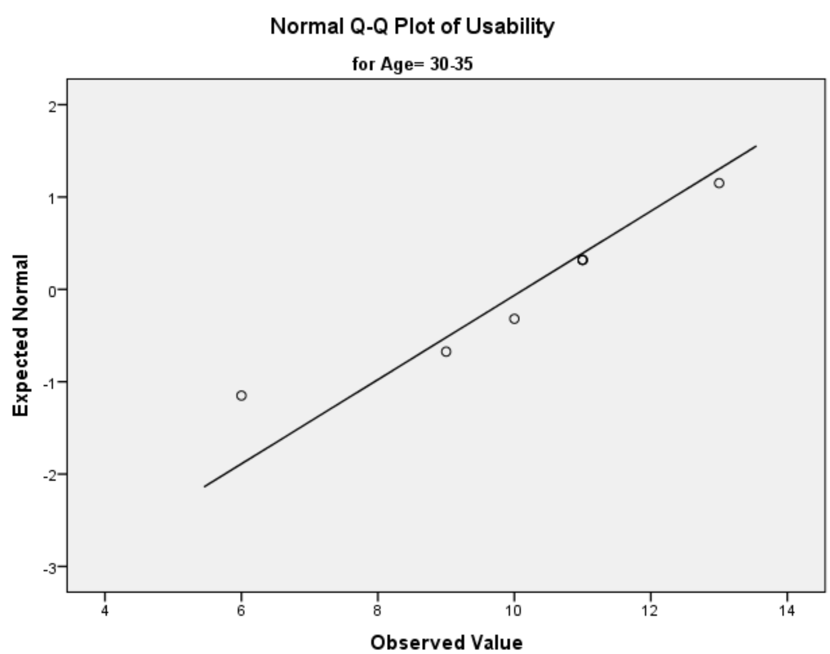


Figure 70. Q-Q plot of usability for age 30-35: Visual expeditionary skills training.

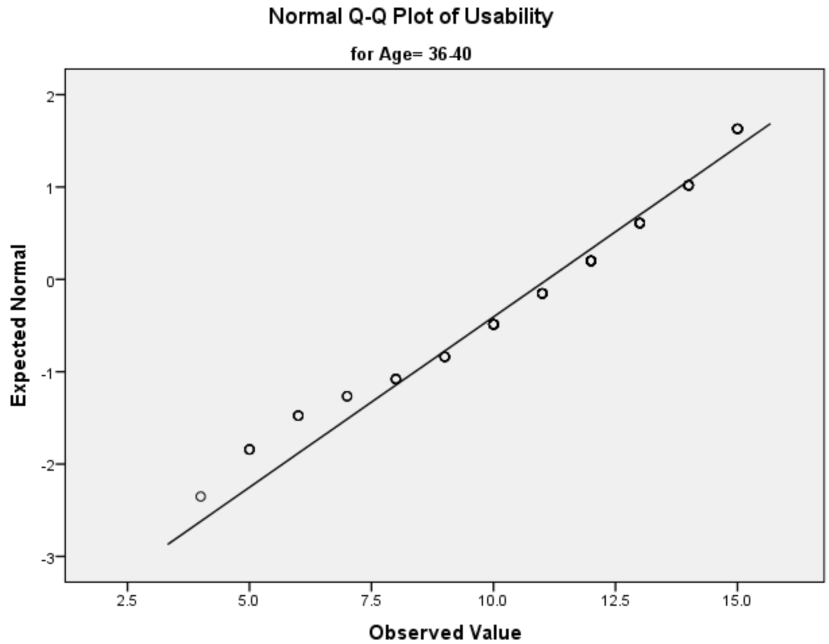


Figure 71. Q-Q plot of usability for age 36-40: Visual expeditionary skills training.

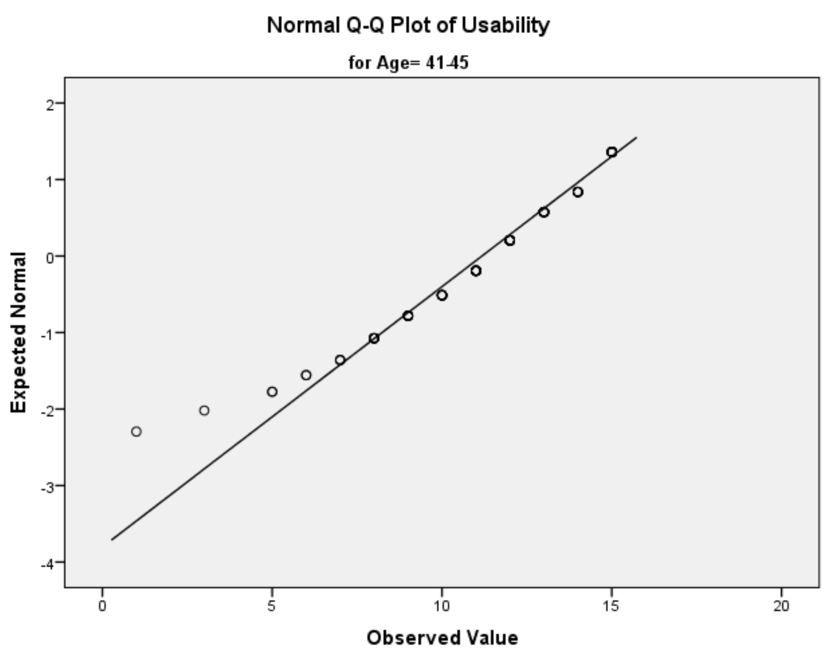


Figure 72. Q-Q plot of usability for age 41-45: Visual expeditionary skills training.

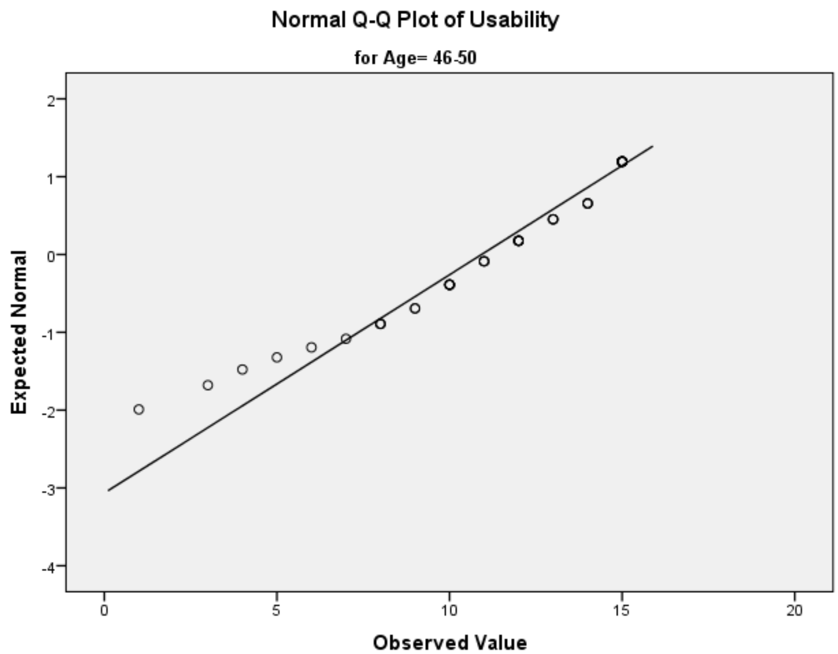


Figure 73. Q-Q plot of usability for age 46-50: Visual expeditionary skills training.

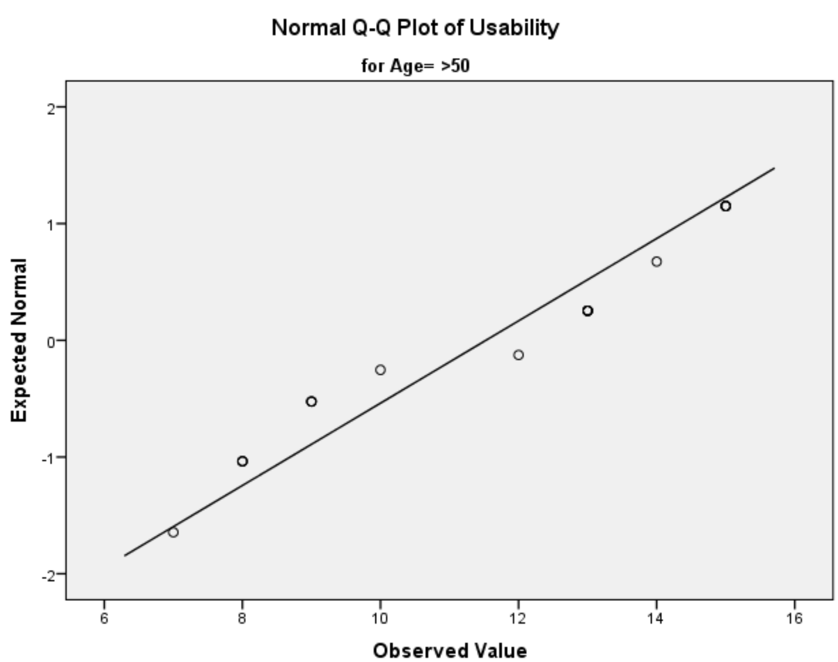


Figure 74. Q-Q plot of usability for age >50: Visual expeditionary skills training.

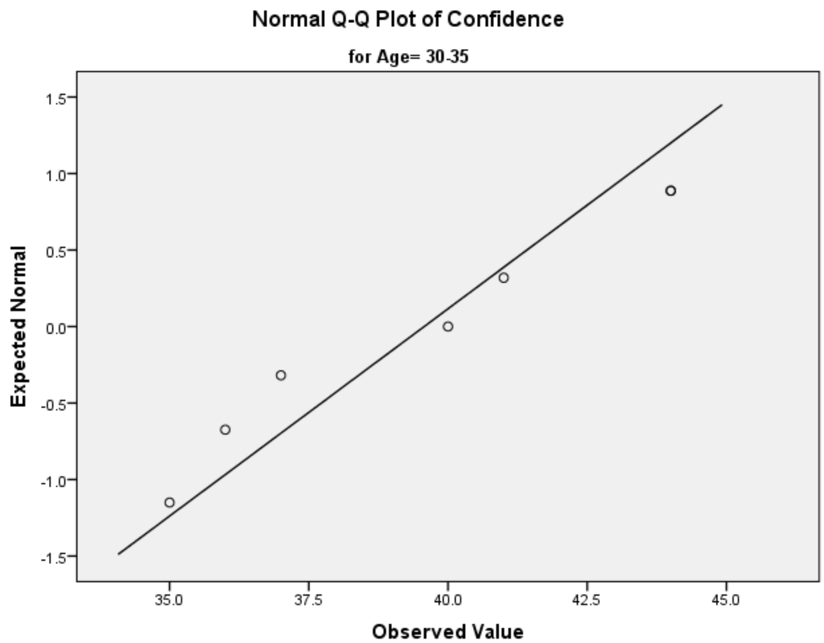


Figure 75. Q-Q plot of confidence for age 30-35: Visual expeditionary skills training.

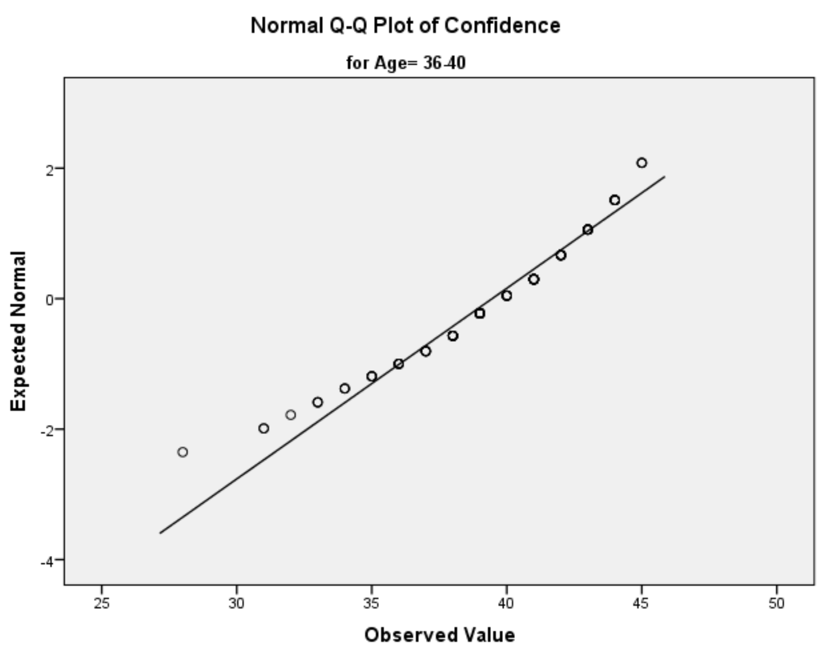


Figure 76. Q-Q plot of confidence for age 36-40: Visual expeditionary skills training.

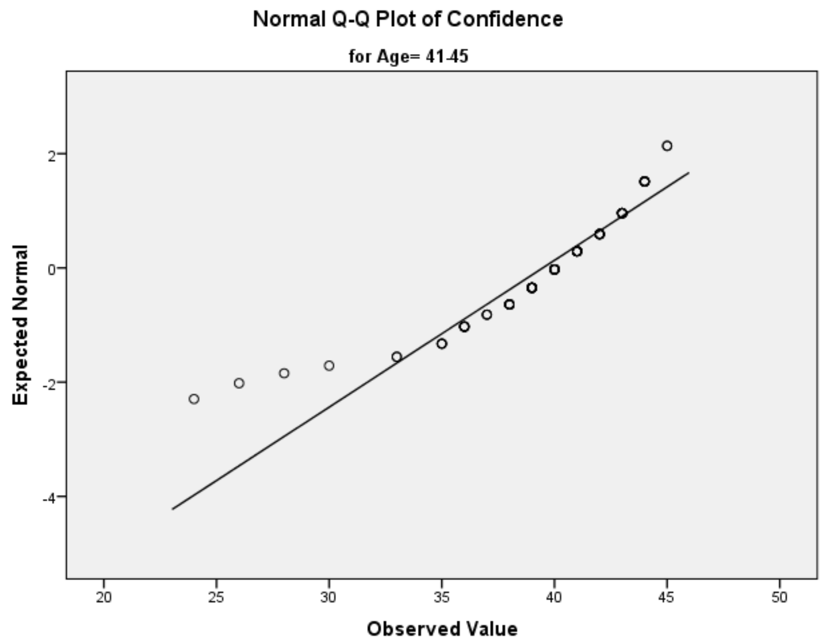


Figure 77. Q-Q plot of confidence for age 41-45: Visual expeditionary skills training.

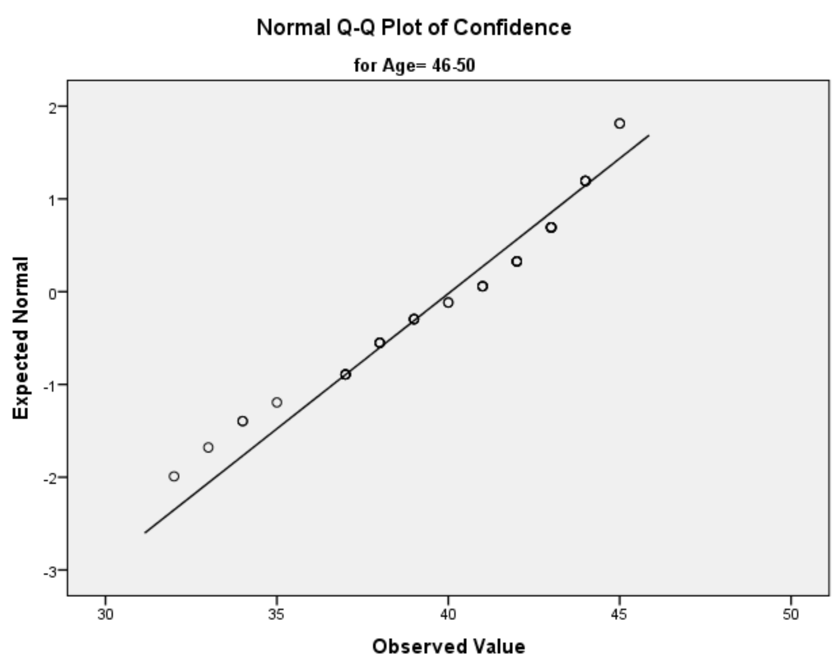


Figure 78. Q-Q plot of confidence for age 46-50: Visual expeditionary skills training.

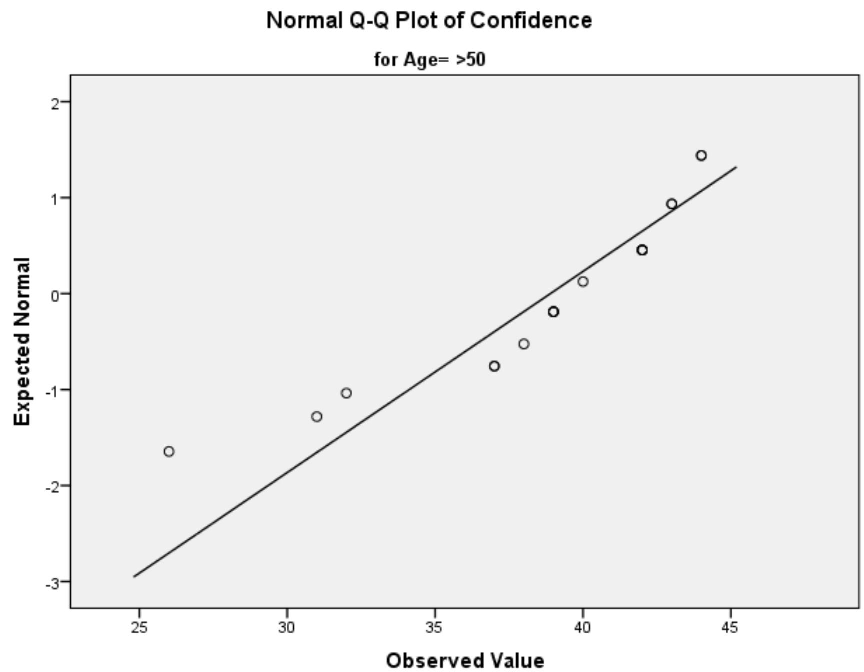


Figure 79. Q-Q plot of confidence for age >50: Visual expeditionary skills training.

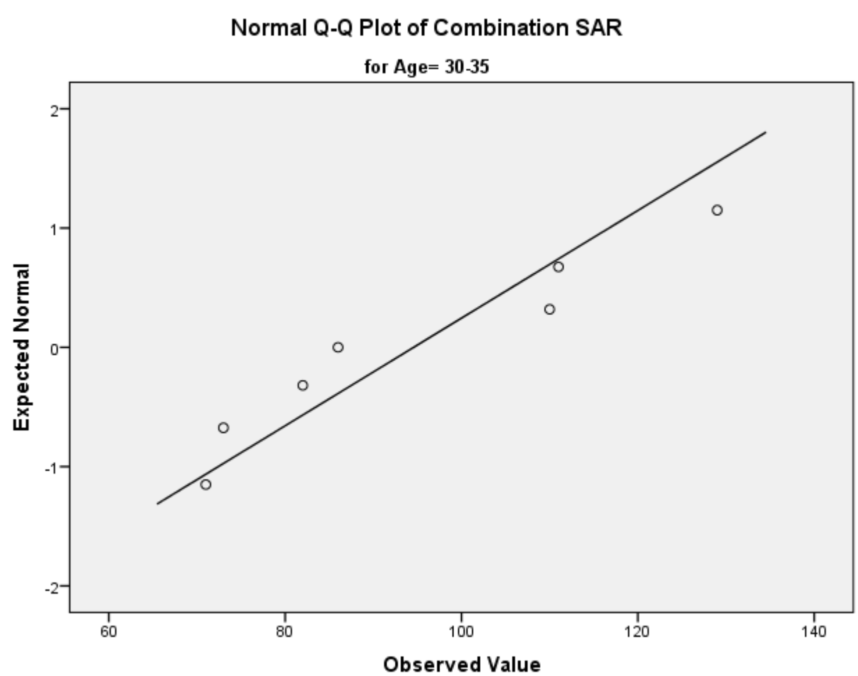


Figure 80. Q-Q plot of combined satisfaction, attention, and relevance for age 30-35: Visual expeditionary skills training.

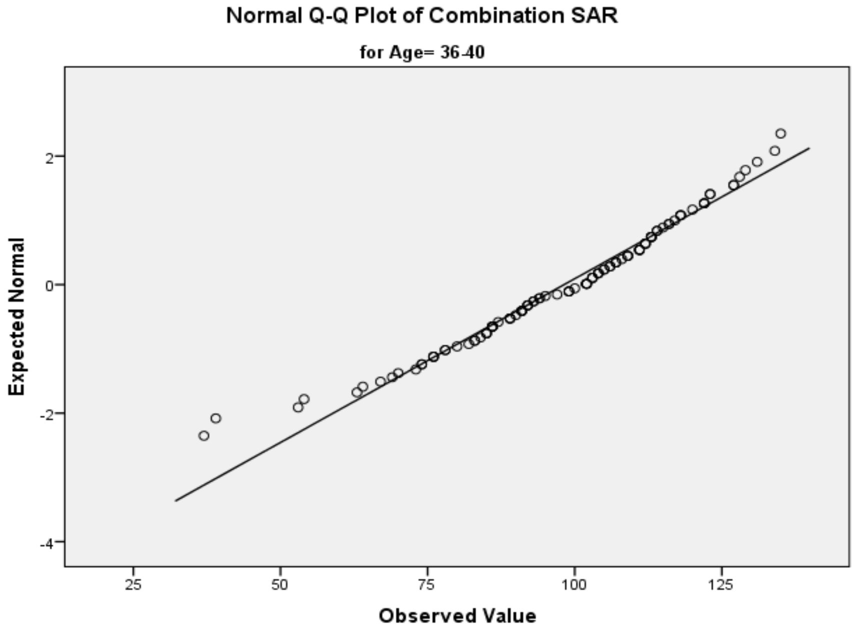


Figure 81. Q-Q plot of combined satisfaction, attention, and relevance for age 36-40: Visual expeditionary skills training.

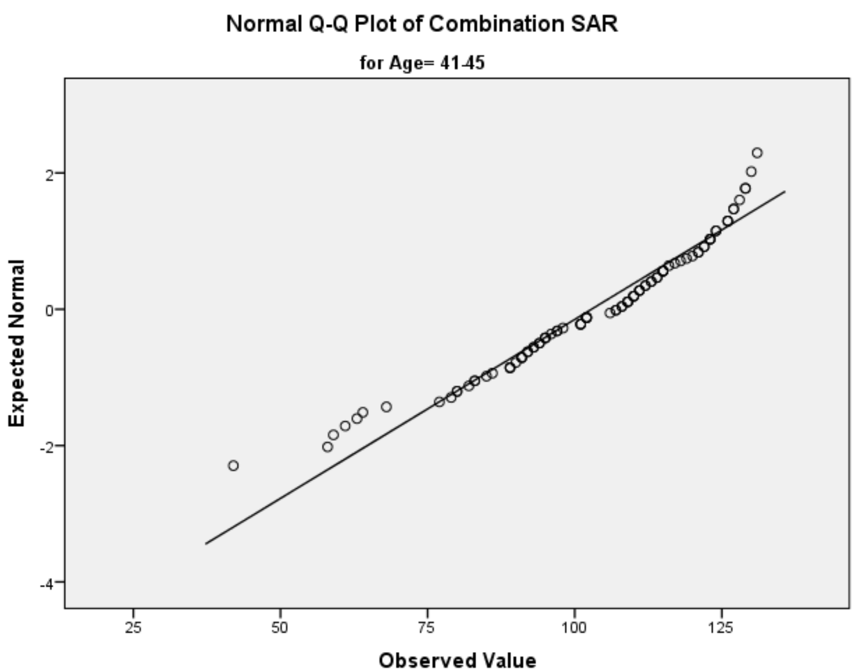


Figure 82. Q-Q plot of combined satisfaction, attention, and relevance for age 41-45: Visual expeditionary skills training.

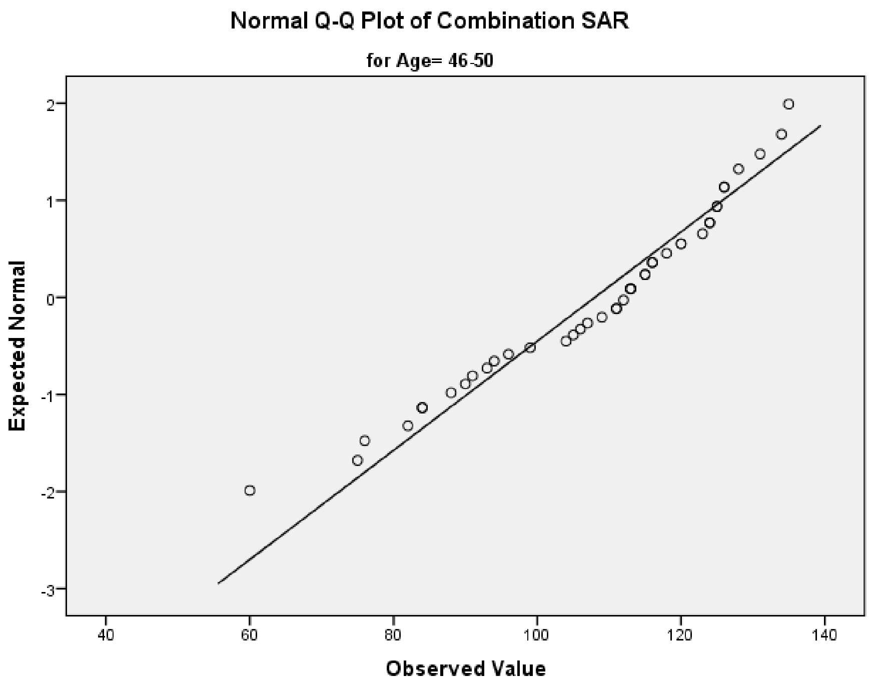


Figure 83. Q-Q plot of combined satisfaction, attention, and relevance for age 46-50: Visual expeditionary skills training.

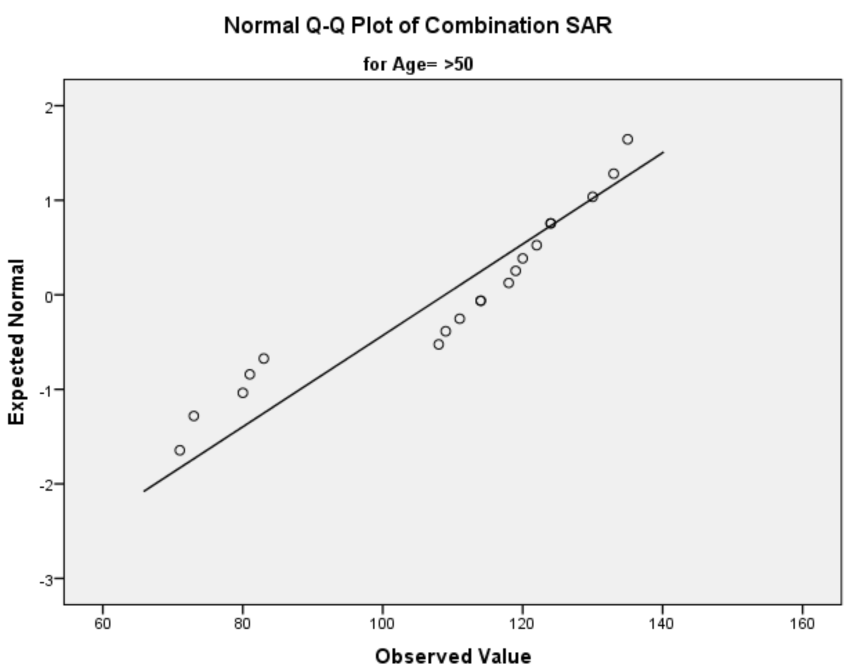


Figure 84. Q-Q plot of combined satisfaction, attention, and relevance for age > 50: Visual expeditionary skills training.

The fifth assumption was linearity, which looked at the relationships between the dependent variables (Lund & Lund, 2013). The scatterplot matrixes shown in Figures 85 through 89 were used to check this assumption. Visually, the scatterplots indicate a linear relationship existed for each of the pairs and therefore the assumption of linearity was met. The least linearity appeared to be between perceived effectiveness and usability, which is similar to the original simulation dataset.

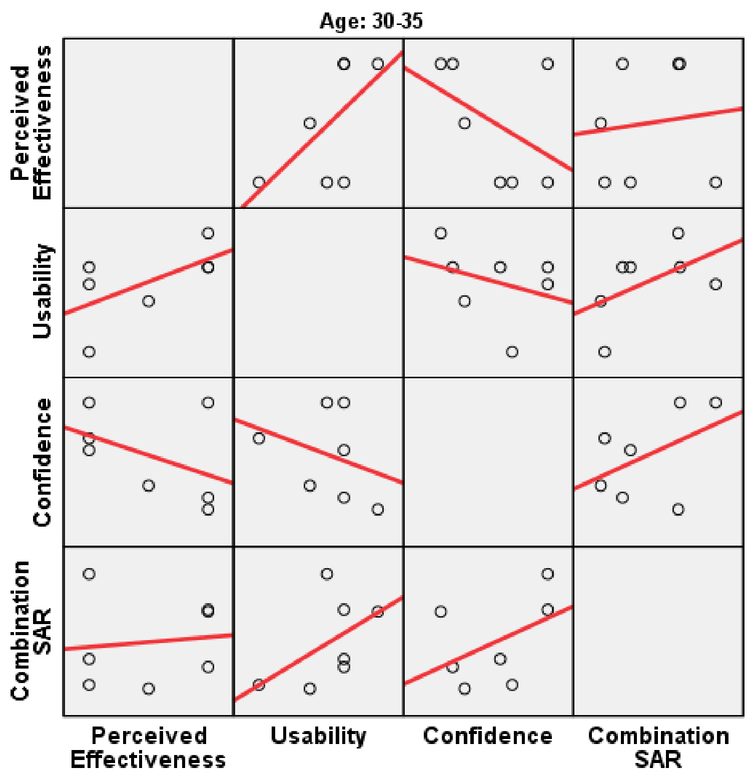


Figure 85. Scatterplot of dependent variables for age 30-35: Visual expeditionary skills training.

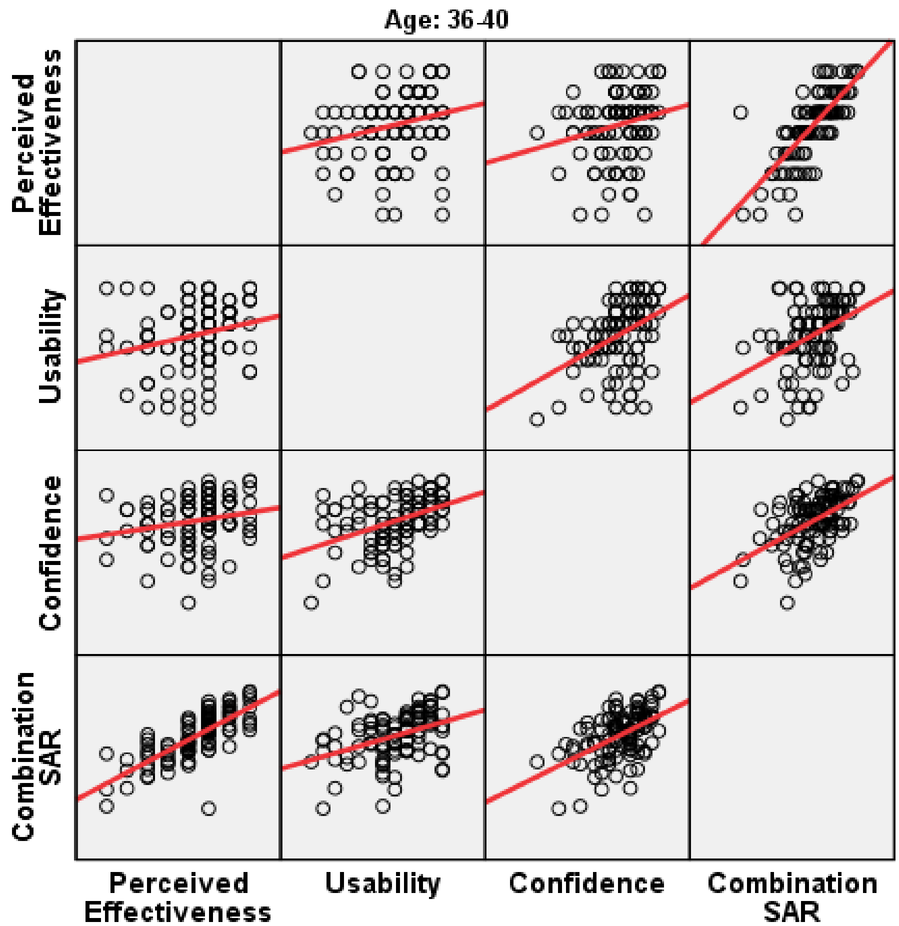


Figure 86. Scatterplot of dependent variables for age 36-40: Visual expeditionary skills training.

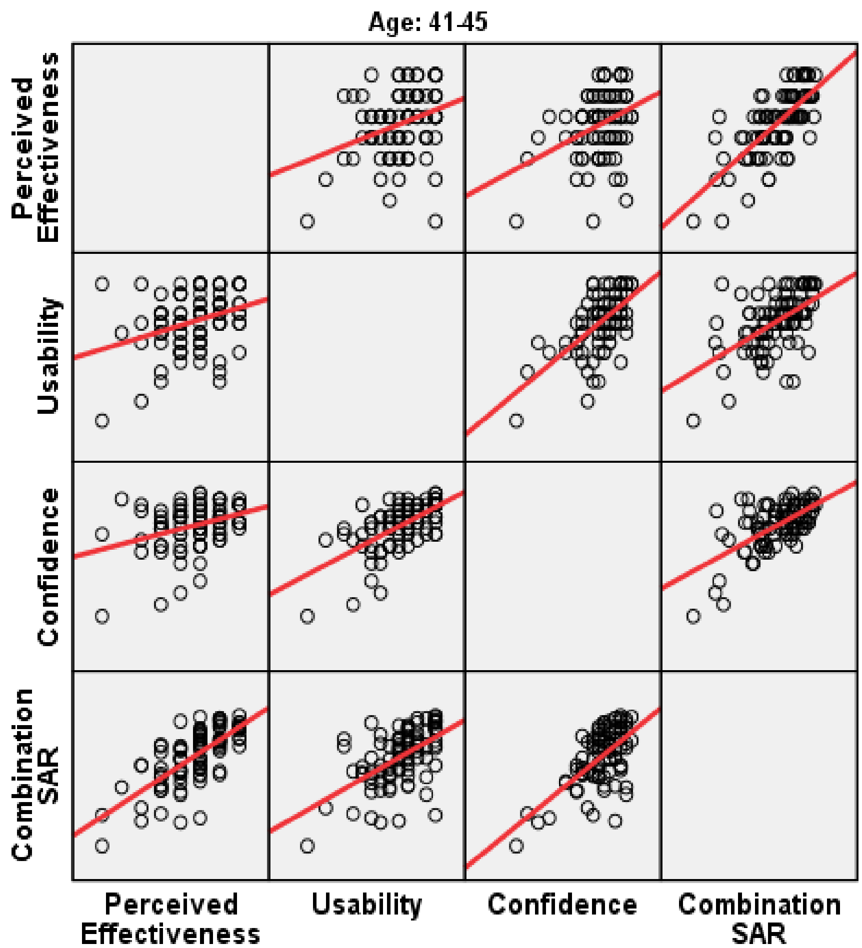


Figure 87. Scatterplot of dependent variables for age 41-45: Visual expeditionary skills training.

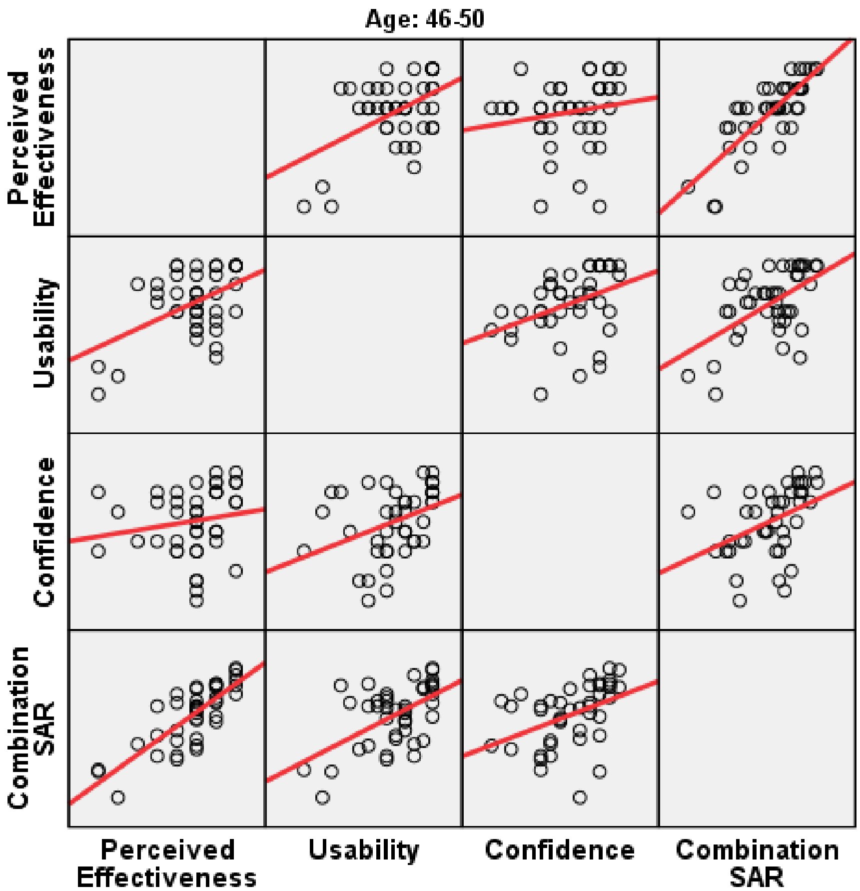


Figure 88. Scatterplot of dependent variables for age 46-50: Visual expeditionary skills training.

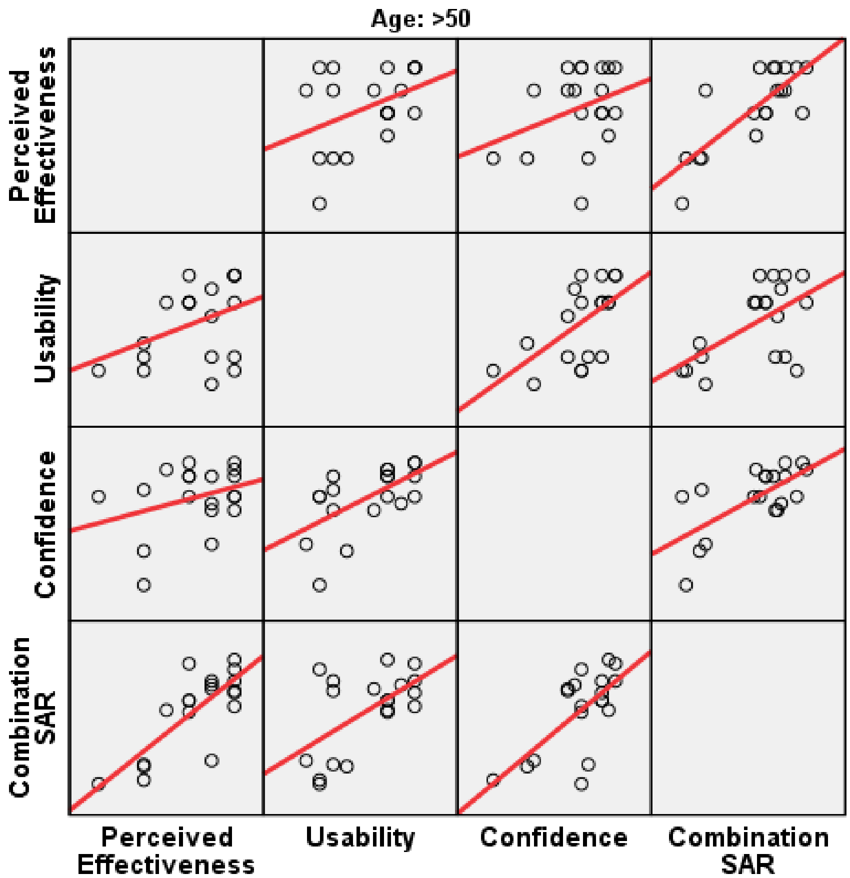


Figure 89. Scatterplot of dependent variables for age >50: Visual expeditionary skills training.

The sixth assumption was the existence of homogeneity of variance and covariances. This was tested using both the Levene's test for individual dependent variables and Box's M for the covariance. Both tests were designed so if significance was found (in this case, less than 0.05), it indicated the sample did not meet the assumption and the variances needed to be treated as unequal. Values of greater than .05 meant the variables met the assumption. The values from the Levene's test for each dependent variable are listed in Table 71. All four variables met the requirements for assuming homogeneity of variances. The check for homogeneity of variances within groups (covariance) was checked using Box's M, which had a *p* value of 0.377, so the null hypothesis was rejected and homogeneity of covariances existed across the groups.

Table 71

Levene's Test of Variances for Each Dependent Variable

Variable	Levene's Test Original Simulation
Perceived Effectiveness	0.810
Usability	0.242
Confidence	0.822
SAR combined	0.863

The seventh and final assumption was no multicollinearity. The dependent variables should be moderately correlated. If they are too low, there is no reason to do the MANOVA and if they are too high (> 0.90), it would indicate multicollinearity, which would be problematic for running a MANOVA (Lund & Lund, 2013). Multicollinearity was checked using Pearson's correlation coefficients between each of the dependent variables. The results are listed in Table 72. All the coefficients indicated moderate correlations with a low of 0.247 for perceived effectiveness and confidence and a high of 0.746 for perceived effectiveness and the combined SAR variable. All pairs showed statistical significance using an alpha of 0.01. Overall, the data met the assumption of no mulitcollinearity.

Table 72

Multicollinearity Check Using Pearson's Coefficients

		Perceived Effectiveness	Usability	Confidence	Combined SAR
Perceived Effectiveness	Pearson Correlation	1	.332	.247	.746
	Sig. (2-tailed)		.000	.000	.000
	N	265	265	265	265
Usability	Pearson Correlation	.332	1	.481	.481
	Sig. (2-tailed)	.000		.000	.000
	N	265	265	265	265
Confidence	Pearson Correlation	.247	.481	1	.546
	Sig. (2-tailed)	.000	.000		.000
	N	265	265	265	265
Combined SAR	Pearson Correlation	.746	.481	.546	1
	Sig. (2-tailed)	.000	.000	.000	
	N	265	265	265	265

Except for outliers, all MANOVA assumptions were met with this dataset as shown in Table 73. There were 49 univariate outliers and three multivariate outliers within 40 samples. Similar to the previous analysis, a MANOVA was conducted with the outliers present and a second one was accomplished with the outliers removed to compare the effect of the outliers on the results.

Table 73

Results of Assumption Check

MANOVA Assumptions	Univariate	Multivariate
1. DV Measured as Interval	✓	✓
2. IV Are Categorical	✓	✓
3. Independent Observations	✓	✓
4. Adequate Sample Size	✓	✓
5. No Outliers	49 identified*	3 identified*
6. Normality	✓	✓
4. Linearity	✓	✓
5. Homogeneity of Variance	✓	✓
6. Multicollinearity	✓	✓

* 40 unique samples contained outliers.