NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

THE NEXT GENERATION OF AVIATION MAINTENANCE TRAINING: GAME-BASED AND VIRTUAL REALITY AUGMENTED LEARNING

by

Enrico L. Landas

September 2018

Thesis Advisor:	Rudolph P. Darken
Co-Advisor:	Perry L. McDowell

Research for this thesis was performed at the MOVES Institute.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** September 2018	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE THE NEXT GENERATION OF AVIATION MAINTENANCE TRAINING: GAME-BASED AND VIRTUAL REALITY AUGMENTED LEARNING	**5. FUNDING NUMBERS**
6. AUTHOR(S) Enrico L. Landas	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Plane captains are an essential part of efficient flight line operations. They coordinate multiple efforts during aircraft ground operations. The current training methods for plane captains have several significant issues.

We built a simple immersive part-task trainer for plane captains using commercial gaming technology. The system consists of two distinct domains: one in which trainees control a virtual aircraft using their body motion and another that provides 360-degree video of common operations. We also conducted an experiment to determine whether such a system improved the performance of plane captains over traditional training methods.

In our experiment, we trained two groups of plane captains: one via traditional methods and the other via the new system. The subjects then performed different signals while a group of subject-matter experts rated them. Although traditionally trained subjects performed better, comments by the graders indicated that the causes of the differences could be fixed in a production version. Additional analysis suggests that those trained with the system are just as confident in their abilities as those trained traditionally.

We also conducted a survey of qualified plane captains who had used the system and the results were overwhelmingly positive.

14. SUBJECT TERMS MOVES, modeling, virtual environments, simulation, training, education, 360 video, Kinect, gesture recognition, aviation training, plane captain, gamification, self-regulated learning, F-35, JSF	**15. NUMBER OF PAGES** 107
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 | Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

THE NEXT GENERATION OF AVIATION MAINTENANCE TRAINING: GAME-BASED AND VIRTUAL REALITY AUGMENTED LEARNING

Enrico L. Landas
Major, United States Marine Corps
BS, University of San Diego, 2007

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MODELING, VIRTUAL ENVIRONMENTS, AND SIMULATION

from the

NAVAL POSTGRADUATE SCHOOL
September 2018

Approved by: Rudolph P. Darken
Advisor

Perry L. McDowell
Co-Advisor

Peter J. Denning
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Plane captains are an essential part of efficient flight line operations. They coordinate multiple efforts during aircraft ground operations. The current training methods for plane captains have several significant issues.

We built a simple immersive part-task trainer for plane captains using commercial gaming technology. The system consists of two distinct domains: one in which trainees control a virtual aircraft using their body motion and another that provides 360-degree video of common operations. We also conducted an experiment to determine whether such a system improved the performance of plane captains over traditional training methods.

In our experiment, we trained two groups of plane captains: one via traditional methods and the other via the new system. The subjects then performed different signals while a group of subject-matter experts rated them. Although traditionally trained subjects performed better, comments by the graders indicated that the causes of the differences could be fixed in a production version. Additional analysis suggests that those trained with the system are just as confident in their abilities as those trained traditionally.

We also conducted a survey of qualified plane captains who had used the system and the results were overwhelmingly positive.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AR	Augmented Reality
ASMT	Aircraft Systems Maintenance Trainer
CBT	Computer-Based Training
COTS	Commercial off-the-Shelf
GOMS	Goals, Operators, Methods, and Selection
ICT	Information and Communication Technologies
NAS	Naval Air Station
SME	Subject-Matter Expert
SRL	Self-Regulated Learning
TELE	Technologically Enhanced Learning Environment
VE	Virtual Environment
VR	Virtual Reality
VMX-1	Marine Operational Test and Evaluation Squadron One

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to extend my deepest gratitude to those who have been instrumental in the production and conduct of this research:

Perry McDowell and Dr. Rudy Darken—thank you for all your guidance throughout this process. I will carry your mentorship well beyond my time at the Naval Postgraduate School.

Dr. Peter Squire and ONR Code 30—your support aided in resourcing vital elements needed to complete this project.

The Battlespace Exploitation of Mixed Reality Lab, specifically Heidi Buck, David Rousseau, and Nelson Cysneros—thank you for lending the hardware and technical expertise needed for all video capture needs.

The Commanding Officer and Marines of Marine Operational Test and Evaluation Squadron One—thank you for your support in the development, organization, and execution of this research.

MOVES FutureTech—your help with application development, technical advice, and instruction were essential to the success of this thesis.

The personnel at the Center for Naval Aviation Technical Training Detachment Eglin—your advice and technical expertise greatly shaped the execution of this project.

Finally, I thank my wife, Ling. Your love and support through long nights of studying and throughout an arduous career has given me much perspective and the drive to do my best.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. PROBLEM DOMAIN

Operating on a military flight line is inherently dangerous, both due to the chaotic nature of the environment and the complexity of the aircraft on hand. The primary personnel responsible for controlling this danger are maintainers designated as “plane captains.” Plane captains are responsible for inspecting the aircraft prior to the pilot coming out, with the pilot before the flight, and after the aircraft returns. These inspections and all the other tasks the plane captain conducts throughout this evolution are immensely important in ensuring the events run smoothly and without incident. Once the pilot is in the aircraft, the plane captain is responsible for communicating the current situation and is the central point of direction for all the supporting personnel and equipment. Non-standard operations and emergencies intensify the importance of the plane captain’s role.

As part of normal flight line operations, plane captains execute various physical movements (“hand-and-arm signals”) during aircraft ground operations to communicate with the pilot when conditions preclude an electronic communications tether. These movements are pivotal in processes such as aircraft movement/handling, startup, refueling, final check (the last inspection of the aircraft prior to departure from the “line”), taxi, and launch/recovery (sending/receiving an aircraft to/from flight). Governed by Naval Air Systems Command document 00-80T-113, the Aircraft Signals Naval Aviation Training and Operating Procedures Standardization Manual (Naval Air Systems Command, 2014), these movements should be consistent across all Naval Aviation platforms. They are made up of over two hundred movements and vary whether conducting day or night operations. Marine and Navy pilots are trained during initial qualification on these signals and their meanings. Precise execution and interpretation of these movements is key for safety and mishap prevention. Ground personnel must have the proper qualifications to control aircraft in this manner.

Naval aviation maintenance personnel attain qualifications via the advanced skills management system, an online tracking system that delineates the specific requirements for each qualification. The syllabus for training plane captains entails the following:

- a series of prerequisite qualifications or licenses
- required readings
- discussions with currently qualified personnel, supervisors, or program managers
- on-the-job training (consisting of multiple repetitions)
- written tests
- practical examinations
- a plane captain selection board, where the board members include supervisors, quality assurance representatives, the squadron aviation safety officer, and the squadron maintenance officer who must decide whether the maintainer has demonstrated the skills, poise, and experience to serve as a plane captain

Despite this wealth of training materials, qualifying plane captains is primarily considered an apprenticeship-based system, the current training consists of methods where qualified personnel pass down practical knowledge to the trainee.

B. RESEARCH PROBLEM AND MOTIVATION

Issues with the current training methods can be found in the areas of standardization, interpretation of text and images, sustainment/proficiency, lack of personnel/aircraft availability to conduct practical examinations/training, and safety.

Techniques degrade as plane captains pass information to trainees who may add to, lose, or misinterpret the information, which we will refer to as "generational drift." Deviation in technique can also occur as trainees misinterpret static images of dynamic actions. Over time, this variability in technique compounded with individual preference or physical limitations when conducting these movements will increase the potential for miscommunication and mishap.

The current training construct compounds the strain on personnel and aircraft. Individual plane captain candidates lack feedback when practicing motions learned from text or video instruction, which can result in negative training transfer. This requires training to be conducted with at least two personnel, one being a qualified plane captain.

There are also safety concerns in the current training system's ability to accurately train emergency procedures. Currently, there is no way to simulate many flight line casualties, such as an engine fire, so most plane captains qualify without having experienced them. This can lead to a lack of proficiency in identifying visual and audible cues during an emergency, which may delay reacting to the situation.

As training systems have improved and become cheaper, high-level leadership has issued guidance recommending their increased use. The Commandant of the Marine Corps desires "more 'reps and sets' by training in simulated environments and embrace experimentation to test and validate new concepts" (Neller, 2017, p. 3). The desire for training improvement is also found in the Navy's ready, relevant learning initiative that "will modernize delivery methods from traditional brick-and-mortar schoolhouses to mobile, multi-media, multi-platform delivery solutions that leverage the best concepts to accelerate learning as individuals, teams and organizations" (Moran, 2016, para. 2). This thesis leverages the rise of virtual reality (VR) technology and game-based learning environments in hopes to provide a basis to address the intent of higher-level leadership.

C. RESEARCH QUESTIONS

The following research questions are the focal point for this thesis:

- Is it feasible to create and utilize a game-based part-task trainer to replace or supplement the current method for training aircraft plane captains? We break this down into three sub-questions: 1) can a trainer built using commercial off the shelf (COTS) technology determine whether a trainee performs hand and arm signals correctly as well as subject-matter experts (SMEs)?; 2) can such a trainer produce trained plane captains as well as the traditional methods of instruction?; 3) Is there a difference between the confidence of plane captains trained in the different methods?

- Will users consider a virtualized game-based learning platform a practical means of increasing sets and reps to attain the qualification of plane captain? We answer this question by surveying experienced plane captains who have used the system.

D. BENEFITS OF STUDY

Conducting research in this domain may result in the capability to address the aforementioned issues of standardization, interpretation of text, sustainment/proficiency, personnel/aircraft availability when conducting practical examination/training, and safety. Incorporating current technology will improve training methods as well as increase opportunities to execute repetitions while learning or sustaining skills. In addition, improvements to the distribution capabilities will afford ancillary personnel the opportunity to learn and practice skills without impeding squadron operations.

Currently, not including any of the pre-requisite reading and licenses needed for this qualification, the syllabus requires the trainee to complete approximately 60 hours of training. Current methods of training double the personnel drain on squadrons, since both a mentor and mentee are required for execution. Additionally, a portion of those hours will also require the use of an aircraft or other assets, making them difficult to plan and schedule. Replacing or supplementing the current training methods with virtualized trainers, aircraft, or assets will result in a drastic reduction in the man-hours and assets needed to attain this qualification. Additionally, these systems' ability to generate emergency scenarios enhances the trainees' depth of knowledge in this area prior to execution or live training, resulting in better trained personnel.

E. SCOPE

This study focused on testing the feasibility of implementation and acceptance of a virtualized plane captain training system. The system utilized a combination of video, graphical representations, gesture recognition with active feedback, and game-based scenarios to represent select aspects of the qualification's syllabus. Testing focused on assessing the standardization and accuracy of hand-and-arm signals as well as the usability of the system. We used a select subset of common movements to represent the majority of the hand-and-arm signals included in the current syllabus as a proof of concept along with a usability study to test the effectiveness of the developed product. This subset included the motions for:

- applying brakes
- opening the canopy
- guiding personnel
- indicating hands off
- conducting a vehicle systems built-in test
- opening the weapons bay doors

F. APPROACH

To address the aforementioned issues, we created a simple immersive game-based part-task trainer utilizing commercial gaming technology, which we named the Plane Captain Training System, which encompasses several aspects of the current training syllabus. This system consists of two distinct domains: one which provides 360-degree video of common operations and another where trainees control a virtual aircraft using their body motion. The system is designed to be split into seven modules that incorporate the two environments throughout. Though not all explored in this thesis, the modules can be described as a progressive learning structure that provides an orientation to the system, indoctrination to the operational environment, an instructional series, practice environments, and proficiency testing with qualification tracking.

G. THESIS STRUCTURE

The remainder of this thesis is structured as follows:

Chapter II outlines a brief history of virtual environments (VEs), discusses some of the current uses of VEs, and ties this technology to the realm of education.

Chapter III describes the task analysis conducted for plane captains controlling aircraft.

Chapter IV outlines the design methodology used and details the development of the Plane Captain Training System prototype.

Chapter V describes the experimentation conducted as well as the corresponding results and analysis.

Chapter VI summarizes the conclusions made and recommendations for future work.

II. LITERATURE REVIEW

A. VIRTUAL ENVIRONMENTS

Throughout recent years, VR technology has been on a steady growth in popularity. The 2017 article "Increasing Adoption of AR and VR in Gaming Expected to Drive Growth of the Global AR and VR Market Through 2025," by PR Newswire, attributed a multitude of factors contribute to this growth in popularity. These factors include games based on augmented reality (AR), an increase in spending on electronic goods for entertainment, the evolution of cell phones, as well as investments to simplify AR and VR technology while at the same time adding more features. PR Newswire projects continued growth of the AR and VR market "which was valued at US$ 5,175 million in 2016" and projected "to reach a valuation of US$ 119,540 million by 2025-end."

Like most new concepts which have not matured sufficiently to have agreed upon lexicon, there are many varying definitions of VR. However, Stanković (2016) gives one of the best, where he defines VR by first discussing the ideal of perception, then delineating between VR and VEs. Stanković describes an individual's perception of reality as a subjective state which "is shaped by two distinct things, our mind and our senses" (Stanković, 2016, p. 22). To delineate between VR and VEs, we can define VR as attempting "to alter one's perception of reality by tricking the senses" through computer-generated stimuli and VEs as providing "the illusion of presence in a place different from one's current physical surrounding" (Stanković, 2016, p. 24). Stanković further enhances these definitions by expounding on the terms interaction and immersion. He outlines how other types of media, such as film or computer aided design software, traditionally feature levels of either interaction or immersion, but rarely both. "The property that sets apart VE from other similar media, such as film and 3D video, is interaction. The user of VE is always able to interact in some way with the environment" (Stanković, 2016, p. 29).

B. CURRENT USES OF VIRTUAL ENVIRONMENTS

1. Industry

a. Sports

Ranging from the fans in the seats or at home to the players on the field, the sports industry can utilize VR in a multitude of ways. Bhardwaj (2018) touches on how VR technologies are changing the way fans consume sporting events. Noting two variations of incorporating VR, Bhardwaj states that "people are really excited at the prospect of being able to wear a headset and feel like you are courtside at a basketball game, or on the sidelines at a football game, all from the comfort of your own home" (para. 1). The other variation takes advantage of augmented reality, for people physically watching at the stadium Bhardwaj mentions that there is "an opportunity to provide headsets that overlay statistics or commentary" (10). Mirt (2017) discusses how the sports industry is "investing millions in VR technology mainly to improve and diversify the training methods of athletes" (para. 2). The aforementioned applications are just a few methods in which VR technology can be incorporated into the sports industry; as VR technology becomes more advanced, so will the applications.

b. Medicine

The medical field is constantly growing; this is in large part due to the evolution of ailments affecting the population, however, advancements in technology are also contributing to this growth and in how we practice medicine. Carson (2015) outlines multiple examples of how VR is influencing the medical field, primarily grouped into three areas: treatments for specified conditions, rehabilitation, and training. Some examples include treatments for conditions such as phobias and anxiety through simulated environments, while other examples utilize VR to enhance pre-existing methods such as the Wisconsin Card Sorting Task to treat brain damage. The article also speaks to the benefits of using VR to provide surgeons another venue to practice procedures or techniques while removing any risk to real patients.

2. Military Aviation

a. Flight Simulators

Considered one of the most dangerous occupations in the world, military aviation encompasses complex aircraft, a multitude of factors needed for sustained flight, as well as multifaceted tactical requirements based on the mission at hand. In the early 1900s, as aircraft started being massed produced, "it quickly became obvious that even the small errors in training of new pilots can lead to catastrophic consequences leading to the loss of human life and considerable material damage" (Stanković, 2016, p. 39). To mitigate these issues, the U.S. army turned to Link Aviation Devices Incorporated to develop the first flight simulators. Shown in Figure 1, these devices were nicknamed "blue boxes" by Army pilots. As aircraft continued to develop in technological advancements, the devices utilized for training and simulation also developed. In the mid-1900s, flight simulators progressed to include "huge 'electronic brains' capable of reproducing any airplane flight with great realism" (Dempewolff, 1954, p. 87). Simulators built by Curtiss-Wright incorporated the nose of a real plane in order to provide flight instrument information to the pilot. This style of flight simulator was said to be "so accurate that it has been used to check the performance of an aircraft before the plane even left the ground" (Dempewolff, 1954, p. 87). This is merely a brief overview of the history of flight simulators; for a more in-depth information, see Stanković (2016).

Today, the Navy is using the latest technology to make aviation training "more cost-efficient, more networked, more high-end and more beneficial to the students" (Eckstein, 2018, para. 4). Figure 2 shows "one of the biggest leaps in capability at NAS Fallon (which) was demonstrated last summer at the CAVE – a dome-shaped Combined Arms Virtual Environment simulator" (Eckstein, 2018, sec. "The Cave"). The use of the CAVE and its capabilities allows pilots to improve performance through increased practice and features allowing the pilot to mimic a wide array of missions.

Figure 1. The “Blue Box” by Link Aviation. Source: Stanković (2016).

Figure 2. The CAVE Simulator Aboard NAS Fallon. Source: Eckstein (2018).

b. Aircraft Systems Maintenance Trainer (ASMT)

As with the complexity of new aircraft and aircraft simulation devices, the training for maintenance personnel has grown beyond traditional classroom instruction. One example of this, utilized at the Center for Naval Aviation Technical Training Detachment Eglin, is the ASMT. The ASMT is a system developed by the AAI Corporation for use by aviation maintainers to learn aircraft systems and the performance of maintenance tasks for the F-35. Figure 3 shows this system, which "delivers powerful training capability in a highly realistic simulated environment" (Tucker, 2009, sec. "High-fidelity Desktop Training & Adaptable Architecture"). The ASMT leverages these virtual environments through self-paced, instructor led, and performance based training that encompasses component localization, removal and installation procedures, and fault isolation.

Figure 3. User Perspective Utilizing the ASMT. Source: Tucker (2009).

C. VIRTUAL ENVIRONMENT USE IN THE CLASSROOM

1. Games / Gamification

Kapp begins an article on games and engagement in learning with an interesting statement that "game-based learning can turn disconnected, bored learners into engaged participants" (2012, p. 64). Kapp examined two case studies that demonstrated some gaps in the "traditional" learning mantras. He outlined the first instance as one that involved a learner attempting to go through an e-learning course where the learner is unable to progress unless they listen to the whole script of each slide; while the script is playing, the learner "checks his email, plays with his smartphone, and absentmindedly clicks when the audio stops" (para. 2). The second scenario is one where the learner is "sitting in the classroom listening to the instructor drone on and on" (para. 3). Kapp states that the problem with these learning styles is that the learners "are not engaged in the learning process" and the "information is presented in a disjointed and unconnected fashion" (para. 4). Kapp contrasts these with learning in places that use games as learning tools and elaborates on the elements of gamification. He states that a "well-designed game is a system in which players engage in an abstract challenge, defined by rules, interactivity, and feedback that result in a quantifiable outcome often eliciting an emotional reaction" (para. 7). Kapp elaborates on the term gamification and states that it "is using game-based mechanics, aesthetics, and game thinking to engage people, motivate action, promote learning, and solve problems" (para. 12). He notes games provide the "freedom to fail" and states, "games overcome the 'sting of failure' by allowing, as part of their design, multiple opportunities to perform a task until mastery" (para. 17). This is relevant to the growth of VR technology and the use of VEs as an increasingly popular venue to incorporate the ideals of gamification.

2. Self-regulated and Technology-Enhanced Learning Environments

Dettori and Persico (2011) present "the relationship between [self-regulated learning] SRL and [information and communication technologies] ICTs from several standpoints, addressing both theoretical and applicative issues, providing examples from a range of disciplinary fields and educational settings" (p. xx). Bernacki, Agular, and Byrnes (2011) review 55 empirical studies related to the areas of technologically enhanced learning environments (TELE) and how they either enhance or detriment the use of self-regulated learning as a tool for acquiring knowledge. The authors interpret the finding of these studies by answering the following questions:

1. What is the theoretical basis for understanding the possible relations among SRL and TELEs?
2. What types of TELE have been used to study these relations?
3. When participants engage in SRL behaviors in a well-designed TELE, do they show greater learning than their peers who engage in fewer SRL behaviors?
4. How have TELEs been shown to promote SRL tendencies in learners?
5. How do pre-existing SRL tendencies influence the ways in which learners interact with TELEs?

They begin with an introduction that discusses the growth and availability of emerging technologies intended for use with SRL and continue to address the proposed questions from specified interpretation methodologies. The authors' "review suggests that TELEs can promote SRL and are best used by those who can self-regulate learning" (Bernacki et al., 2011). The authors have outlined multiple criteria for when self-regulation may be required; these are as follows:

- The environment is focused on complex, multi-step tasks in which possible solution strategies and outcomes are not known in advance (so the learner must plan and monitor performance).

- It is easy for the learner to become distracted, lose interest, or forget the main goals of the task.
- The task requires the use of strategies (e.g., note-taking) to overcome the processing limitations of the mind.
- Learners must engage in helpful behaviors (e.g., planning, monitoring, strategy use, etc.) on their own, without guidance, pressure, or prompting from others.

The aforementioned emerging technologies include the use of VR and virtual environments and have strong ties to the criteria outlined for when self-regulation may be required.

3. Experiential Learning

Kolb defines experiential learning “as a particular form of learning from life experience; often contrasted with lecture and classroom learning” (2014, p. xviii). Kolb provides a short synopsis of the history of experiential learning and comments on the role that the evolution of experiential learning plays in “shaping and guiding the development of the new educational programs based on experiential learning” (p. 4). The examples given previously in the discussion of current uses of VR show how VR is being used to create experiential learning situations. Because VR technology changes the way we are able to experience the world, it will allow learners to have significantly more and varied experiences than they could otherwise encounter. Therefore, it will have a profound effect on the theory of experiential learning. Bricken states, “text, oral, and screen-based presentations address subsets of human capacity. In contrast, the VR learning environment is a context that includes the multiple nature of human intelligence: verbal/linguistic, logical/mathematical, auditory, spatial, kinesthetic, interpersonal and intrapersonal” (1991, p. 179). Research into how we can incorporate VR technology into the learning process will deepen our ability to leverage this growing phenomenon.

III. TASK ANALYSIS

A. INTRODUCTION

We conducted a task analysis in order to complete an in-depth system design of the training device. A task analysis is "the study of what an operator (or team of operators) is required to do, in terms of actions and/or cognitive processes, to achieve a system goal" (Kirwan & Ainsworth, 1992, p. 1). In this case, the operators are the plane captains and the system goal is safely conducting all necessary actions to launch an aircraft. This study of actions and cognitive processes drove the system design so it can appropriately relay and assess the information.

There are many varying types of task analysis; Kirwan and Ainsworth divide a representative set of these task analysis methods into the following groups:

- Task data collection methods
- Task description methods
- Task simulation methods
- Task behavior assessment methods
- Task requirements evaluation methods

Considering the system goal of conducting the necessary actions to launch an aircraft and the desired functionality of the proposed application, the methodology for task analysis will potentially touch on all these methods.

B. TASK ANALYSIS METHODOLOGY

Helander, Landauer, and Prabhu state "that empirical user testing is too slow and expensive for modern software development practice, especially when difficult-to-get domain experts are the target user group" (Helander, Landauer, & Prabhu 1997, p. 733). When choosing a task analysis methodology, the performer must consider the scope, prototypical nature, and timeline of the study. In this case, testing focused on assessing the standardization and accuracy of hand-and-arm signals as well as the usability of the system. We chose to utilize the goals, operators, methods, and selection (GOMS) model to conduct the task analysis after an exploration of various other techniques failed to capture the complexity and hierarchical nature of these movements and associated prerequisites. “A GOMS model is a description of the knowledge that a user must have in order to carry out tasks on a device or system; it is a representation of the ‘how to do it’ knowledge that is required by a system in order to get the intended tasks accomplished” (Helander et al., 1997, p. 734). As a representative set of hand-and-arm signals, we have de-constructed the “engine start” and “engine fire” signals as representative of all the signals we implemented. Identified as having an elevated risk to safety, personnel, and assets, these specific movements require meticulous training due to their complexity in execution.

In order to de-construct the movements, we assembled a team to conduct a “walk-through-talk-through” as covered in (Ciavarelli, 2017). This team included an aircraft maintenance officer as the SME and one other non-aviation member, referred to as team member two. The SME on the team conducted the movements while observed by team member two. Following completion of the movement and observation by team member two, the SME team member then completed the signaling movement again, this time verbalizing their actions and cognitive processes. Team member two recorded the actions in the GOMS format.

The team members then conducted a review of the recorded signals. The SME re-read the GOMS outline to team member two. Team member two completed the movements following only the verbal instructions read by the SME. The team then noted and remedied any deficiencies between the predicted and actual signal execution.

C. GOMS RESULTS

Figure 4 is a sample of the task analysis for the start engine hand-and-arm signal; Appendix A. Task Analysis Results displays the full analysis of the two representative movements.

Assumptions:

(1) Individuals have successfully passed basic military training.
(2) Individuals have completed primary military occupational specialty school as an aviation maintenance technician.
(3) The task is being completed as a part of normal garrison squadron operations.
(4) Post/before operations servicing inspections have been completed the night before.

Limitations:

(1) This task does not include ancillary checks for persistent maintenance issues attributed to specific aircraft.

Goal 1. Start Engine
Method for Goal 1.
    Goal 1.1 Prepare aircraft
    Method for Goal 1.1
        Operator: Identify and carry appropriate tools (motor)
        Operator: Identify and don appropriate personal protective equipment PPE (motor)
        Operator: Walk out to identified aircraft (motor)
        Operator: Visually inspect aircraft for abnormalities (perceptual)
        Operator: Stow red gear pins/covers/plugs (motor)
        Operator: Conduct debrief with pilot (cognitive)
        Operator: Close ladder bay door (motor)

    Goal 1.2 Position/Posture for engine start up
    Method for Goal 1.2
        Operator: Walk 15ft to the left of the nose of the aircraft (motor)
        Operator: Walk 2ft forward of the nose of the aircraft (motor)
        Operator: Turn to face aircraft and pilot (motor)
        Operator: Assume parade rest (motor)
        Operator: Confirm pilot signal integrated power pack start (perceptual)
        Operator: Confirm integrated power package exhaust is clear (cognitive)
        Operator: Reciprocate pilot signal (motor)
        Operator: Assume parade rest (motor)
        Operator: Confirm integrated power package start (perceptual)

Figure 4. A Sample of the Task Analysis for the Start Engine Hand-and-Arm Signal

The results of the task analysis aided in shaping desired system design elements. Some key items discovered throughout the task analysis needed to optimize the system design included:

- Modules to help identify the appropriate tools needed
- Modules to help identify the appropriate personal protective equipment

- A representation of the needed asset or aircraft (to include sound)
- A system to gauge or track the user's distance from the asset
- A representation of the pilot or pilot cues to the plane captain
- Modules to help teach the necessary hand-and-arm signals
- A need for gesture recognition to assess precession of movements

IV. SYSTEM DESIGN AND DEVELOPMENT

A. DESIGN METHODOLOGY

The system design assessment evaluated proposed system architectures to determine the appropriate system and media for the transfer of information to the plane captain trainees. Along with the recommendations derived from the task analysis, the system needed to successfully convey information to the plane captain trainees in such a way they could easily understand and learn it, and they could be evaluated in accordance with the task analysis. For the design assessment, we considered two proposed system architectures, referred to as solution A and solution B, outlined below.

1. Solution A

Solution A consisted of an interactive game-based system utilizing graphical and video representations to provide initial environment orientation as well as demonstrate procedures and movements. This system utilized commonly used gesture recognition software and technology (i.e., the Microsoft Kinect) to provide feedback during practical exercises and examinations as well as potentially interface with current qualification tracking systems to provide updates on progress and examinations. The following describes the seven modules of the envisioned learning system:

- Module 1 consisted of an orientation to the system. Turning on the system would prompt the user for an optional overview of the user interface, navigation, and synopsis of module contents.

- Module 2 was a guided exploratory module which would provide a 360-degree video with a senior plane captain conducting an aircraft launch. It would display all needed actions from work center preparations, maintenance meeting snippets, and a standard aircraft launch from the flight line. This module was intended to provide the plane captain trainee an opportunity to gain contextual awareness of what plane captain duties entail.

- Module 3 was designed as the first in the instructional series where hand and arm signal instruction would be provided; the user would also be provided the opportunity to practice with active feedback. Movements would be demonstrated in stages (walk-through/talk-through) via video capture prior to affording the trainee the opportunity to imitate the gestures with immediate and active feedback, where the system would give the trainee a percentage of accuracy and instruction on how to correct deficiencies. Deficiency correction would be provided via graphical representation of degrees off from desired action.

- Module 4 incorporated a graphical representation of the aircraft that would respond to the actions provided by the trainee to gain association of hand and arm signals to the output of the aircraft (visual and auditory). This module was organized as a part task trainer, meaning the user will be able to choose any set of actions to practice. Active feedback would still be in play and the aircraft would not respond unless the user is within a threshold of accuracy.

- Module 5 provided the trainee a procedural set or scenario-based environment. The trainee would need to run all procedures and signals but would not be afforded active feedback. The trainee would only be successful or unsuccessful in launching the aircraft and given accuracy scores once the scenario is complete.

- Module 6 incorporated the capability for emergency actions as a result of incorrect signal from the trainee or mechanical malfunction. Scenarios would be capable of being corrected provided the trainee executes the appropriate corrective actions.

- Module 7 would interface with the qualification tracking system and administer knowledge and practical examinations.

2. Solution B

Solution B consisted of a standardized computer-based training (CBT) program. The program would be accessible from any government computer or individually owned laptop through a link provided to plane captain trainees. The CBT would contain four sequential modules that would be locked, to be completed in order. This system would not replace formal evaluation but would provide a means with which trainees may practice without formal interaction from supervisors. This system was divided into three modules that are describe as follows:

- Module 1 reviewed pre-flight checklist items to be completed prior to engine start. This module included the assumption that normal maintenance, post and before operation servicing, was complete. The pre-flight would guide trainees through the required pre-flight items, with a description and accompanying pictures. There would be several requirements in Module 1 for the student to identify, items such as the location of panels, critical inspection areas, and pin storage locations.

- Module 2 would review the procedures and necessary actions prior to aircraft launch. The CBT would show, in sequential order, the movements of the plane captain with visual descriptions. Imbedded in Module 2 would be a recording of a fully qualified plane captain conducting the movements in the correct order. Module 2 would break down the movements according to the task analysis.

- Module 3 would perform the evaluation. This module would be a scheduled Skype call with a centralized evaluation call center. Skype software would be utilized for the trainee to video chat with an identified experienced plane captain. The evaluator would be required to hold a plane captain evaluator qualification rating. During the Skype call, the evaluator would provide a scenario description to the trainee for which they would have to respond appropriately. The trainee would then complete the required verbal description of the activities and complete the hand-and-arm signal for the evaluator. The evaluator would not provide feedback during this portion. The evaluator would upload remarks into the centralized training database then debrief the student on recommendations for improvement and noted deficiencies.

3. Pros and Cons of Proposed Solutions

a. Solution A

(1) Pros

- Utilizing this system will improve standardization as there will be a central repository for the graphics, videos, and scenarios provided.
- All training materials need to only be created once, which will reduce the amount of development personnel. This also reduces any deviation due to differences in instructors' methods.
- Graphical representations can present emergency procedures without putting personnel or material assets at risk.
- Active feedback allows for immediate correction or incorrect movement. This prevents trainees from retaining incorrect information.
- System and scenario updates can be done via network connections.

(2) Cons

- Additional equipment will be required for operation. Equipment will need to be included on service schedules.
- The number of participants is limited to the number of purchased and available systems.
- Additional training of personnel will be required for use and familiarization of various interface mediums.

- Initial material or front-end graphical/video material creation is anticipated to be time consuming and costly.

b. Solution B

(1) Pros

- Time to build training infrastructure is minimal. Requirements for CBT is minimal, internet connectivity and a laptop.

- Equipment/resource requirements are minimal. Most trainees own a laptop with Skype or video capabilities. Additional video resources (monitor mounted camera) may be purchased by the squadron for use by trainees. The expected cost for this additional piece of equipment is minimal.

- As system B is a centralized CBT program, updates can be instituted quickly and with minimal system downtime.

(2) Cons

- System B requires dedicated evaluators. The evaluators may be at a geographically separate location but need to be available for trainees to conduct their Skype evaluation. Additional personnel for computer support may be required.

- Government computers may be able access the CBT but may have connectivity limitations for the Skype evaluation portion.

4. Solution Determination

Evaluation measures were derived utilizing specified constraints supplemented by a few general measures. Each measure was then assigned a threshold and objective parameter as seen in Table 1.

Table 1. System Design Threshold / Objective Metrics

Evaluation Measure	Threshold	Objective	Optimal Metric
Cost	5	10	**low cost**
Modular	3	10	**ability to start/stop/resume individualized training**
Ease of use	7	10	**accommodate minimal technical knowledge**
Throughput	7	10	**can run continuously**
Network Access	8	10	**ability to receive updates via network**
Interoperability	4	10	**ability to connect to ASM**
Size	8	10	**no need for dedicated space**
Lifecycle Maintainability	6	10	**sustainable with minimal cost beyond initial investment**

Note: These parameters are kept in a rating scale of 1 to 10, with 10 being the most optimal.

Each proposed system was then assessed and scored according to the constraints provided. These scores can bee seen in Table 2.

Table 2. Proposed System Constraint Scoring

Raw Data Matrix		
Evaluation Measure	**Alternative 1**	**Alternative 2**
Cost	8	9
Modular	9	9
Ease of use	8	8
Throughput	9	7
Network Access	9	9
Interoperability	9	7
Size	8	9
Lifecycle Maintainability	8	9

Applying the Clemen and Reilly method (Parnell, Bresnick, Tani, & Johnson, 2013), we calculated swing weights, shown in Table 7 of Appendix B. System Design Supplementals. We then created a decision matrix utilizing the raw score of each attribute to make a utility score, which is then multiplied by the weight and summed, seen in Table 3.

Table 3. Proposed System Decision Matrix

Decision Matrix							
	Raw		Utility				
Evaluation Measure	Alt 1	Alt 2	Alt 1	Alt 2	Weight	Alt1 x Wt	Alt2 x Wt
Cost	8	9	0.8	0.9	16.39%	0.13	0.15
Modular (Ref. constraint 2.1.1)	9	9	0.9	0.9	14.75%	0.13	0.13
Ease of use (Ref. constraint 2.1.2)	8	8	0.8	0.8	14.75%	0.12	0.12
Throughput (Ref. constraint 2.1.3)	9	7	0.9	0.7	13.11%	0.12	0.09
Network Access (Ref. constraint 2.2)	9	9	0.9	0.9	9.84%	0.09	0.09
Interoperability (Ref. constraint 2.3.1)	9	7	0.9	0.7	9.84%	0.09	0.07
Size (Ref. constraint 2.4.1)	8	9	0.8	0.9	11.48%	0.09	0.10
Lifecycle Maintainability	8	9	0.8	0.9	9.84%	0.08	0.09
					Sum =	0.85	0.84

With a weighted score of 0.85, solution A was selected for prototype development.

B. PROTOTYPE DEVELOPMENT

Due to limitations on time and the prototypical nature of the development of the Plane Captain Training System for the purposes of this research, only modified versions of modules 2–5 were created. The architecture to house each module was created using Unity, a commercial off-the-shelf (COTS) game engine.

For the 360-degree videos in module 2, we used multiple Samsung Gear 360s to capture a standard aircraft launch for use. We filmed highly experienced plane captains at the Marine Operational Test and Evaluation Squadron One (VMX-1) F-35 Detachment aboard Edwards Air Force Base to create the aircraft launch and instructional videos. The F-35 Joint Program Office reviewed the videos to ensure that they did not include any classified information, and the Battlespace Exploitation of Mixed Reality Lab at Space and Naval Warfare Systems Center Pacific produced the final video.

For modules 3–5, we used the Microsoft Kinect's gesture building software development kit and the Unity game engine to create the applications.

Shown in Figure 5 and Figure 6, qualified plane captains traveled to the Naval Postgraduate School to aid in the gesture recognition development video capture. The Naval Postgraduate School's software and game development team, Future Tech, created the game architecture and graphical representations of assets/personnel.

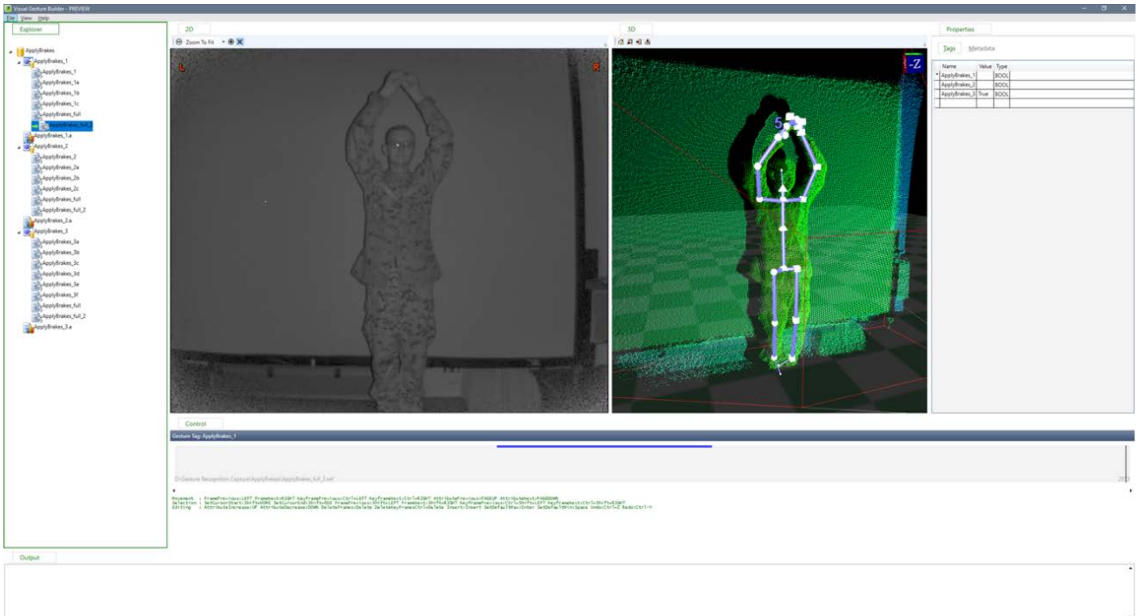


Figure 5. Capturing Gestures by a Qualified Plane Captain from VMX-1 for the Plane Captain Training System

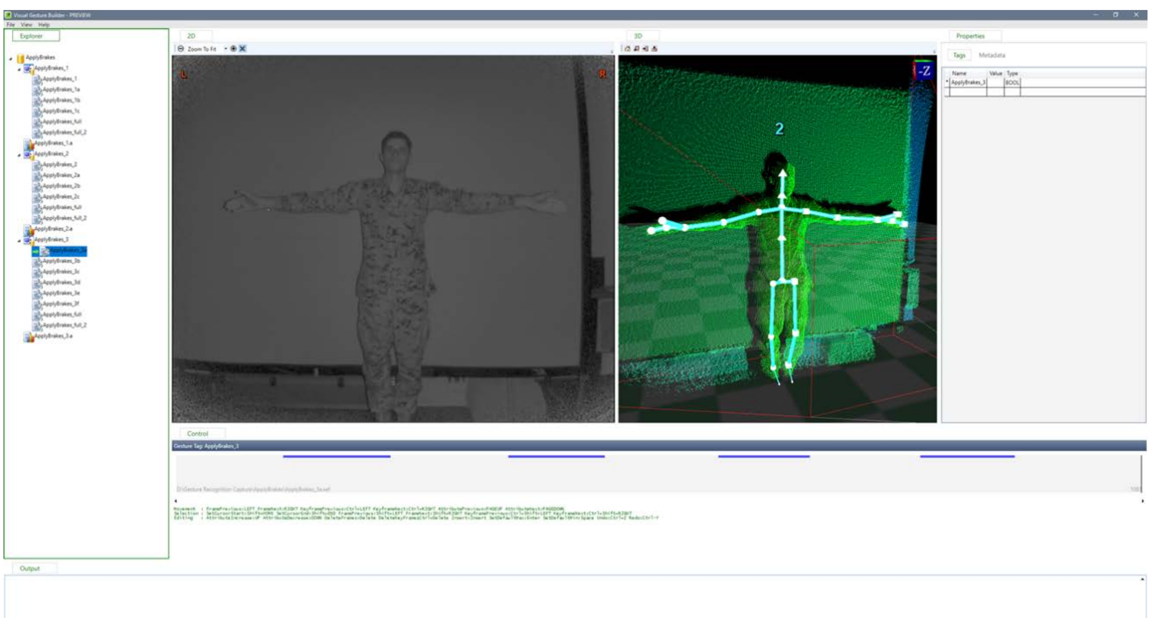


Figure 6. Capturing Gestures by a Qualified Plane Captain from VMX-1 for the Plane Captain Training System

1. User Interface

When beginning to use the system for training, the plane captain trainee comes across the main menu (Figure 7), which includes options for the following:

- Kinect Calibration: this menu item guides the user through steps to ensure the Microsoft Kinect is functioning properly.
- The Plane Captain Experience: this menu item consists of module 2, the exploratory module.
- Gesture Trainer: this menu item consists of modules 3 and 4, affording the trainee the opportunity to learn and practice hand-and-arm signals.
- Scenario: this menu item consists of module 5, allowing the trainee to conduct all the necessary steps to launch an aircraft.
- Freeplay: affords the trainee the opportunity to execute any desired movement for practice.
- Exit: this item will exit the game.
- Log In: this item will afford the trainee the opportunity to enter their name and keep track of individual progress.

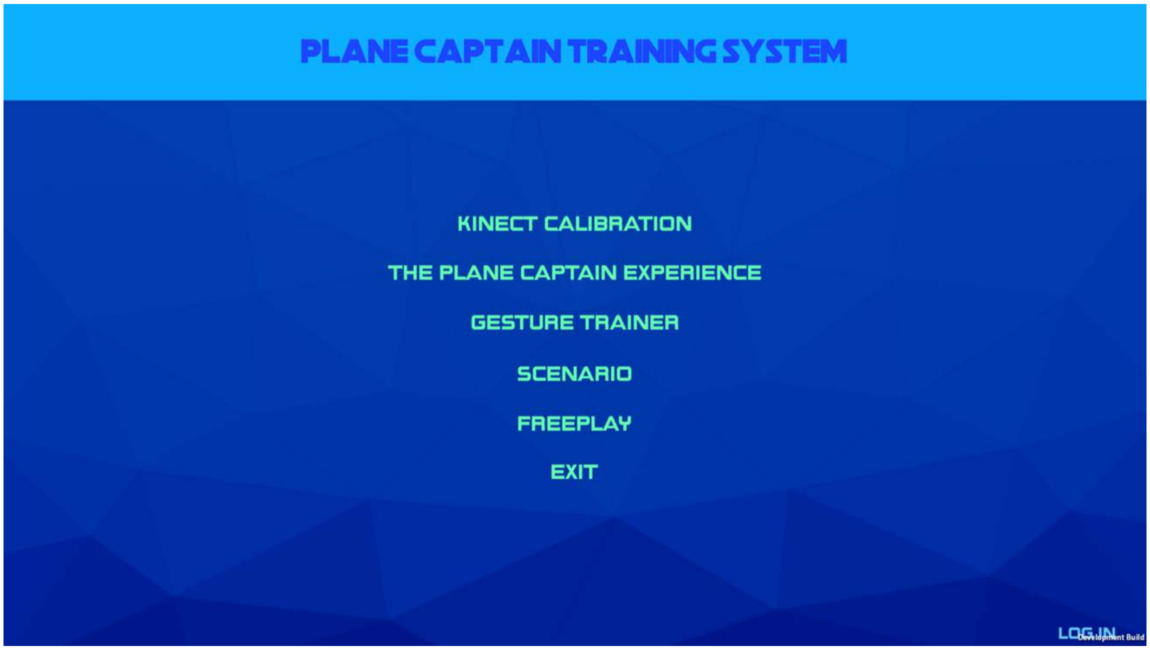


Figure 7. The Plane Captain Training System Main Menu

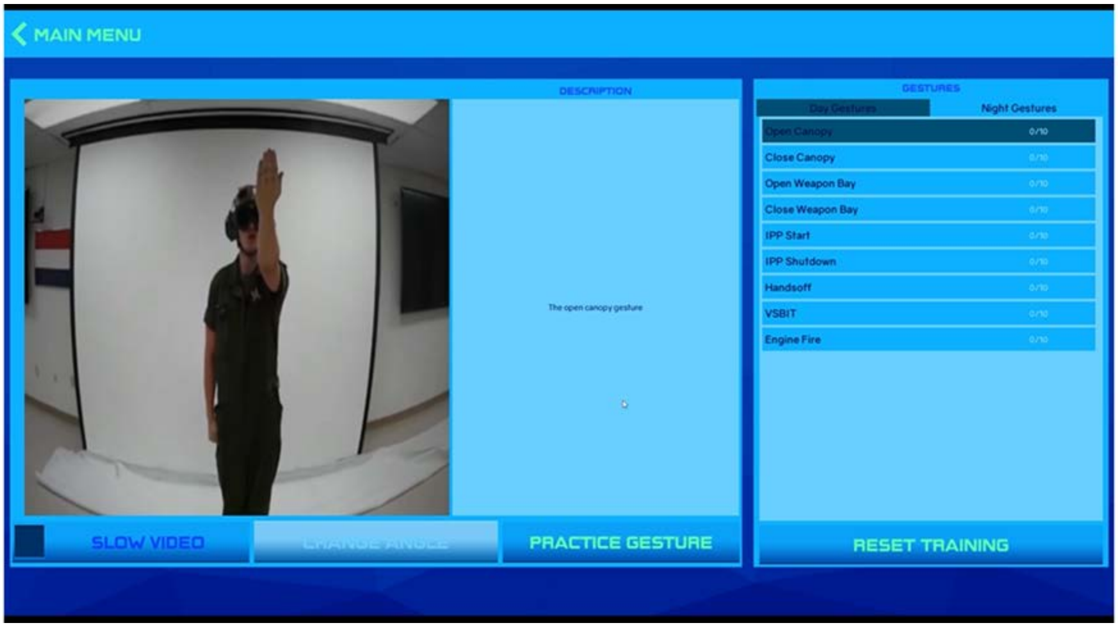
2. Exploratory Module

The exploratory module, shown in Figure 8, allows the plane captain trainee to select between five locations to observe 360 video of an experienced plane captain launching an aircraft. The trainee is also able to click and drag on the screen to change the viewing perspective, i.e., look in different directions from the same location. Additionally, the plane captain in the video is verbally explaining all the actions they are taking. This module provides the trainee the overall context for each of the movements they will encounter in the instructional series.

Figure 8. Plane Captain Training System Module 2 Interface

3. Instructional Series

The instructional series is comprised of module 3, shown in Figure 9, and module 4, shown in Figure 11. The trainee is able to select any hand-and-arm signal from the list and will be able to watch as an experienced plane captain demonstrate the gesture while explaining how the selected hand-and-arm signal should be conducted. The trainee has the option to also slow the video or change the angle of the video during playback. Once the trainee is comfortable with the selected signal, they can practice the gesture while observed by the Kinect. The system compares the trainee’s movements to those required to correctly complete the signal and provides feedback. A deviation from the original design of these modules is that the trainee is not given a percentage of accuracy and instruction on how to correct deficiencies. Due to time and programming limitations, the trainee is relegated to only being given an indication that the movement was conducted either correctly or incorrectly via a counter, shown in Figure 10. The trainee is also able to display the associated video during practice if needed. Once the trainee has successfully completed the movement ten times, they will be able to continue on to module four and practice with a graphical representation of the aircraft, shown in Figure 11. If the trainee correctly completes the movement, the system will direct the model to perform the directed action. If the movement is completed incorrectly during this module, the model will not react.



Note: This image is early in the development stage and does not display all the hand-and-arm signals available in the version used during experimentation.

Figure 9. Plane Captain Training System Module 3 Interface

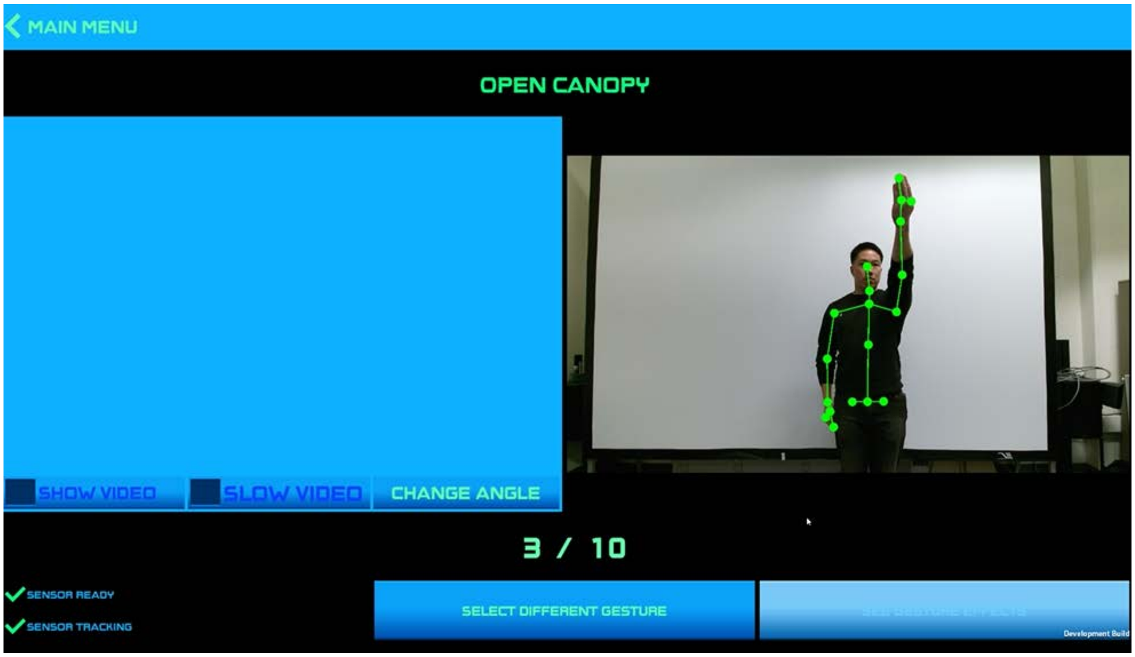


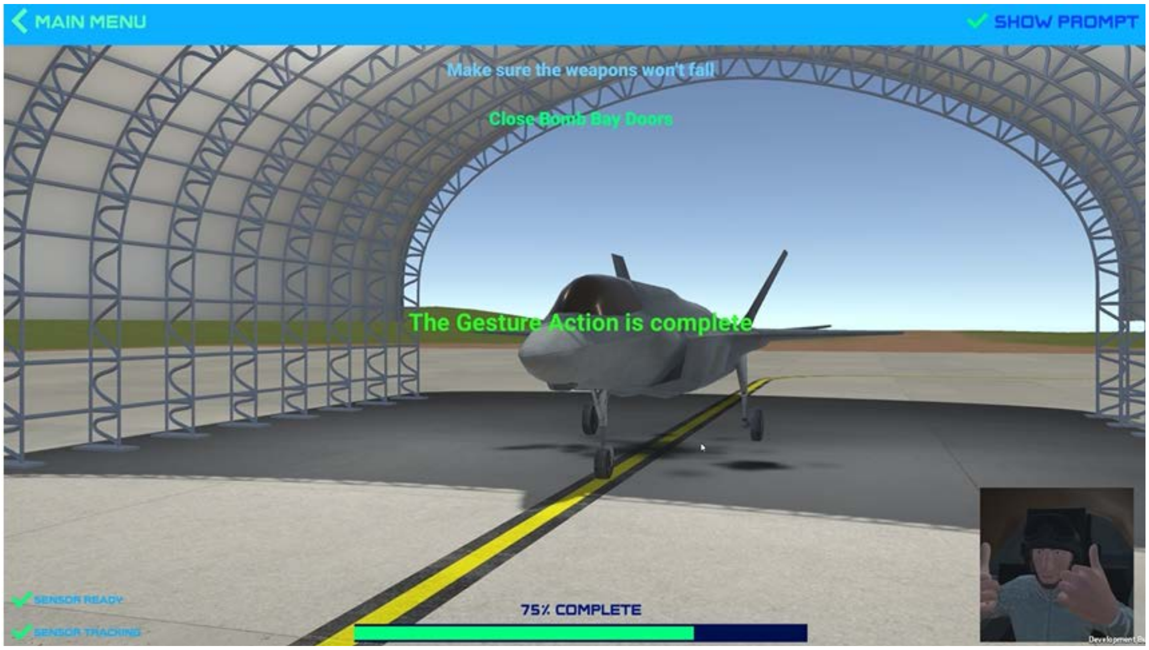
Figure 10. Practicing in the Plane Captain Training System Module 3

Note: A red "X" next to "sensor tracking" indicated that the trainee has left the sensor's field of view.

Figure 11. Plane Captain Training System Module 4 Interface after Successful Performance of Hand-and-Arm Signal

4. Procedural Series

The initial scenario provided to the trainee in module 5 is a traditional aircraft launch. The trainee has the option to show prompts for the expected next movement or they are able to base their movements solely on audible cues or the cues given by the animated pilot. A progress bar is displayed at the bottom of the screen to show aid the trainee is approximating where they are in the launch process. An example of this module is shown in Figure 12.



Note: This snapshot is early in the development stage and is used for conceptualization of this particular module.

Figure 12. Plane Captain Training System Module 5 Interface

V. EXPERIMENTATION AND ANALYSIS OF RESULTS

A. EXPERIMENTAL DESIGN

To assess the developed prototype, we conducted an analysis designed with two main components. The first component was an experiment to answer our first research question (Is it feasible to create and utilize a game-based part-task trainer to replace or supplement the current method for training aircraft plane captains?) and its two sub-questions (Can a trainer built using COTS technology determine whether a trainee performs hand and arm signals correctly as well as SMEs? Can such a trainer produce trained plane captains as well as the traditional methods of instruction?).

The second component was a survey of experienced plane captains to answer our second research question: Will users consider a virtualized game-based learning platform a practical means of increasing sets and reps to attain the qualification of plane captain?

1. Subjects

Subjects in both the experiment and the survey were Marine enlisted personnel who worked as aircraft maintainers. The target population for the experiment participants (n = 25) were personnel that have not received any formal plane captain training. The target population for the usability study participants (n = 11) were current plane captains or personnel that have in-depth knowledge of the functions of the plane captain, such as quality assurance representatives.

2. Location

The experiment was conducted at the VMX-1's F-35 Detachment aboard Edwards Air Force Base. This particular detachment was chosen due to VMX-1's mission given in the 2018 Marine Aviation Plan to "support further concept development and refinement of Marine aviation tactics, techniques, and procedures (TTP)" (Headquarters Marine Corps, 2018, p. 148) as well as the complex nature of the F-35. The personnel at VMX-1 provide a wide variety of experience due to the mixture of initial accession maintainers, meaning the F-35 is their first platform, and maintainers that have previously worked on other type/model/series of aircraft.

3. Conducting the Experiment

After receiving a short overview brief and providing informed consent, all participants completed a demographic survey, shown in Figure 26 of Appendix C. Experiment Supplementals, to collect relevant background information. The participants were assigned to either the usability study or the experiment based on whether they had ever been qualified as a plane captain or had extensive experience on the flight line.

The experiment participants were randomly split into two groups. The first group, referred to as the traditional group, received training via the current method (n = 12). An experienced group of plane captains taught the subjects select hand-and-arm signals using the current text/apprenticeship approach for a given amount of time. The second group (n = 13) received training on the same hand-and-arm signals to learn using the Plane Captain Training System for the same amount of time. All the subjects then demonstrated the movements in front of a review board consisting of the Kinect, which leveraged the developed gesture recognition, and who annotated a pass or fail grade on the scoring sheet shown in Figure 28 of Appendix C. Experiment Supplementals. We collected data on how the Kinect and SMEs graded each gestures that trainees performed.

We allowed the usability study participants to utilize the Plane Captain Training System for a given amount of time, after which they were asked to complete the usability survey. The usability study consisted of having qualified plane captains complete a User Interface Usability Evaluation questionnaire modified from the Gary Perlman web-based questionnaires (Perlman, 2015, sec. “USE”), shown in Figure 27 of Appendix C. Experiment Supplementals.

4. Hypotheses

Throughout each component of this analysis, we will compare the following:

- The average pass rates of the graders during the post task scoring.

- The mean confidence level as annotated by the participants post task.

- For the usability study, we compared the mean rating of the reliability coefficient (Cronbach's Alpha (α)) for the usability study. The results section explains this coefficient in further detail.

a. Experiment—Post-task Scoring

(1) H_0: $P_K = P_{SME}$

(2) H_a: $P_K \neq P_{SME}$

Note: P_K and P_{SME} are the average pass rates of the Kinect and the SMEs.

b. Experiment – Post-Task Confidence

(1) H_0: $\mu_T = \mu_G$

(2) H_a: $\mu_T \neq \mu_G$

Note: μ_T is the mean confidence level of the traditional group and μ_G is the mean confidence level of the group utilizing the game application.

c. Usability Study—Comparison of Mean

(1) H_0: $\mu_{C(x)} \geq 3.00$

(2) H_a: $\mu_{C(x)} < 3.00$

Note: $\mu_{C(x)}$ is the mean user response corresponding to each construct outlined in the usability study (Usefulness, Ease of Use, Ease of Learning, and Satisfaction). The value of 3.00 denotes the midpoint response available to the users from a 1 to 5 scale.

d. Usability Study—Determination of Reliability

(1) H_0: $\alpha \geq 0.70$

(2) H_a: $\alpha < 0.70$

Note: α is the reliability coefficient as defined by Cronbach's Alpha and 0.70 is the acceptable reliability coefficient as defined by Nunnally (1978).

B. RESULTS

1. Usability Study

Eleven total subjects, ranging from the ranks of E-3 to E-8, 3 to 28 years of service, and 21 to 47 years of age, participated in the usability survey. These subjects were all either current plane captains or had previously qualified as a plane captain and indicated that they were confident in their ability to perform the functions required of the job.

Table 8, found in Appendix D. Usability Study Analysis Results, displays the answers provided by the subjects participating in the usability survey with four color coded underlying constructs that include Usefulness, Ease of Use, Ease of Learning, and Satisfaction. Subjects were asked to rate the provided statement from a scale of -2 to 2, corresponding to a range of “disagree” to “agree”; these responses were transposed to a 1 to 5 scale for analysis purposes. Subjects were also afforded the opportunity to provide positive/negative comments, which were categorized and used in the discussion section. Survey items and scale can be referenced in Figure 27 of Appendix C. Experiment Supplementals.

Table 4 displays the average response values off all the usability study participants for each item of the survey, as well as the overall average for each underlying construct. Items 3, 8, and 27 have been excluded from these averages and will be expounded on in the reliability of survey data section. Overall averages without exclusion can be found in Table 9 of Appendix D. Usability Study Analysis Results.

Table 4. Usability Study Total Average User Response (Modified)

Underlying Construct	Item Number						Overall Average
Usefulness	Q1	Q2	Q4	Q5	Q6	Q7	μc(Usefulness)
Overall Average User Response	4.18	4.36	3.91	4.36	4.18	3.90	4.15

Underlying Construct	Item Number											Overall Average
Ease of Use	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	μc(Ease of Use)
Overall Average User Response	4.45	4.55	4.64	4.18	4.09	4.00	4.09	3.73	4.27	4.18	4.30	4.23

Underlying Construct	Item Number				Overall Average
Ease of Learning	Q20	Q21	Q22	Q23	μc(Ease of Learning)
Overall Average User Response	4.45	4.55	4.64	4.73	4.59

Underlying Construct	Item Number						Overall Average
Satisfaction	Q24	Q25	Q26	Q28	Q29	Q30	μc(Satisfaction)
Overall Average User Response	3.91	4.55	4.73	4.00	3.70	4.36	4.21

Note: Average user responses are calculated with items 3, 8, and 27 removed.

After administering a multiple-item rating scale, such as the usability survey utilized for this study, we want to determine the internal consistency of responses to the scale by calculating Cronbach's Alpha. "Cronbach's Alpha is a general formula for scale reliability based on internal consistency. It gives the lowest estimate of reliability that can be expected for an instrument" (Lehman, O'Rourke, Hatcher, & Stepanski, 2013, sec. "Cronbach's Alpha"). In this case, the scale is the -2 to 2 rating format, reliability is defined "in terms of the consistency of the scores obtained on the observed variable," (Lehman, O'Rourke, Hatcher, & Stepanski, 2013, sec. "Reliability Defined") and "instrument" refers to the Plane Captain Training System usability survey.

Each underlying construct was evaluated for Cronbach's Alpha, shown in Figure 30 through Figure 35 of Appendix D. Usability Study Analysis Results. Constructs that were not in an acceptable range, meaning that the reliability coefficients were less than 0.70 (Nunnally, 1978), were re-evaluated with redundant or erroneous items removed. This was based on the rule of thumb "that if an item α is greater than the overall α that includes the item, then scale reliability improves when that item is dropped from the set" (Lehman et al., 2013, sec. "Cronbach's Alpha"). Re-evaluation led to the removal of items 3, 8, and

27 from their perspective constructs and will be discussed in a later section. The results of the remaining items are shown in Table 5.

Table 5. Cronbach's Alpha

Underlying Construct	Cronbach's Alpha	Cronbach's Alpha (Standardized)
Usefulness	0.7598	0.7748
Ease of Use	0.7085	0.7332
Ease of Learning	0.8219	0.8153
Satisfaction	0.7151	0.7267

Note: Analysis of Cronbach's Alpha were conducted with items 3, 8, and 27 removed.

2. Post-task Scoring

After training, all subjects demonstrated each gesture ten times in front of three SMEs while the Kinect also evaluated each gesture. The SMEs rated whether the subject passed or failed each iteration using the sheet found in Figure 28 of Appendix C. Experiment Supplementals, along with providing any comments to explain their. Scoring with the Kinect was conducted in a similar manner where a pass or fail was annotated whether the gesture recognition software felt the subject demonstrated the task correctly.

We utilizing recursive partitioning, also known as partition trees, to analyze the results of the post tasks scoring by the Kinect and the SMEs. These partition trees are a non-parametric approach to describing the relationships between inputs. The partition platform recursively partitions data, automatically splitting the data at optimum points (Sall, Stephens, Lehman, & Loring, 2017). Because of this method, each split is statistically significant, i.e., the differences in values in the two subtrees are not due to chance. The 600 data points from the average Kinect and SMEs scores for the 25 participants conducting the 6 tasks were split 7 times and can be referenced in Figure 36 of Appendix E. Post-Task Scoring Analysis Results.

The first split delineated between the Kinect and the SMEs, shown in Figure 13, showing there was a difference in the scoring between the Kinect and SMEs.

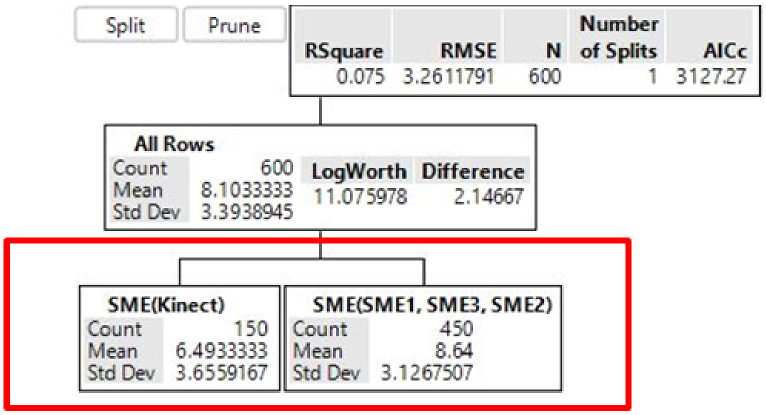


Figure 13. Split #1 (inside Red Box) of Post-task Scoring Partition Tree

Split #2 in Figure 14 shows that, by the SMEs grading, the plane captains trained by the traditional methods outperformed those trained via the Kinect. Split #5 in Figure 15, shows that the Kinect made the same evaluation.

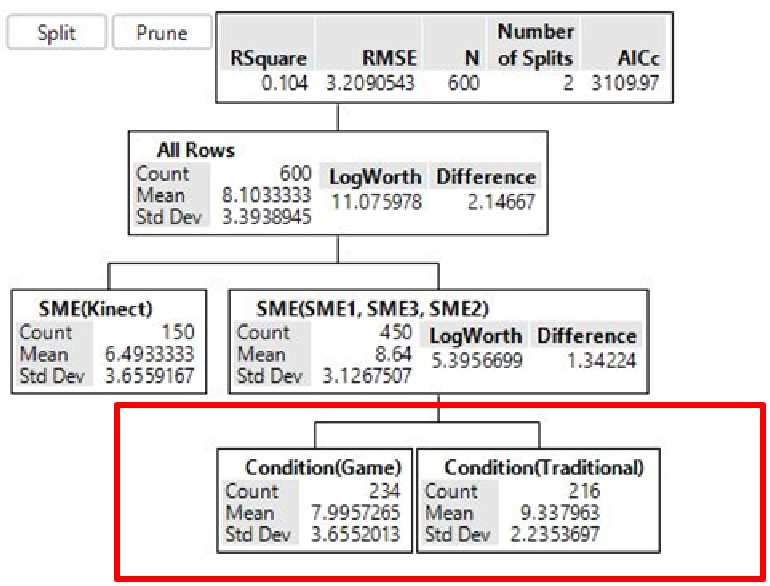


Figure 14. Split #2 (inside Red Box) of Post-task Scoring Partition Tree

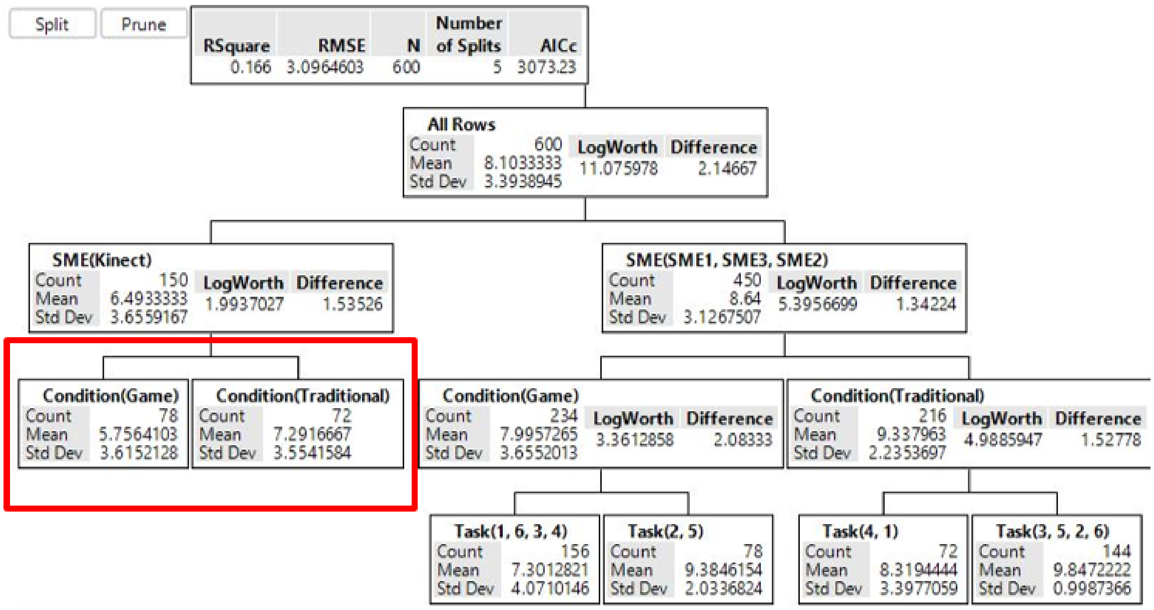


Figure 15. Split #5 (inside Red Box) of Post-task Scoring Partition Tree

Figure 16 shows splits #3, #4, #6, and #7, which further broke the game and traditional conditions into specified tasks. At this point, we stopped breaking nodes apart, as we met our criteria to stop and wanted to avoid overfitting.

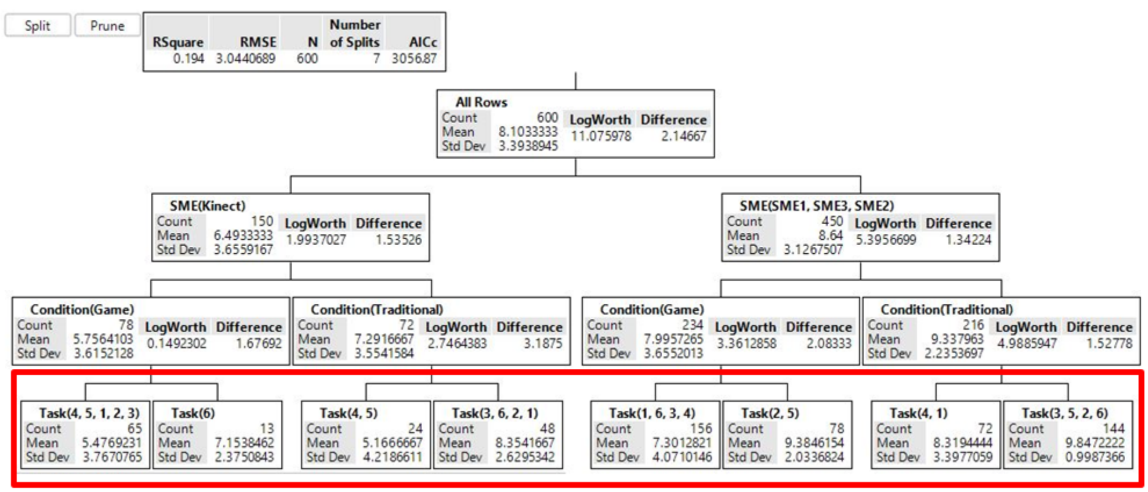


Figure 16. Splits #3, #4, #6, and #7 (inside Red Box) of Post-task Scoring Partition Tree

3. Post-task Confidence

Each subject was given the opportunity to annotate how confident they were in reproducing the material after their training session. Table 6 displays each subject's post task confidence level. Subject participating with the traditional group are highlighted in yellow and the subjects participating with the game application are highlighted in blue. The baseline annotation for "not confident" is 0 and 67.06 for "extremely confident."

Table 6. Post-task Grading Results

Subject ID #	1. Apply Brakes	2. Open Canopy	3. Guiding Personnel	4. Hands Off	5. VSBIT	6. Weapons Bay Open	Overall Average
1	57.02	55.46	55.04	55.21	56.12	56.62	55.91
2	33.55	41.10	14.76	42.70	39.19	43.71	35.84
3	61.50	61.50	61.21	59.37	60.00	62.44	61.00
5	64.84	65.77	66.13	65.27	65.23	65.31	65.43
6	62.77	65.00	63.93	64.50	65.05	65.25	64.42
7	44.15	49.10	43.49	38.37	48.80	48.33	45.37
8	65.24	65.24	65.77	65.67	65.52	65.52	65.49
9	64.87	63.48	4.70	65.75	64.74	64.84	54.73
10	65.55	65.65	32.15	65.69	65.52	65.79	60.06
11	67.06	67.06	67.06	67.06	67.06	67.06	67.06
13	63.79	62.43	63.80	63.39	59.87	63.81	62.85
20	54.55	58.48	54.17	54.17	59.37	60.08	56.80
21	67.06	67.06	67.06	67.06	67.06	67.06	67.06
22	67.06	67.06	67.06	67.06	67.06	67.06	67.06
23	35.27	40.43	24.68	41.04	40.35	43.60	37.56
24	43.38	54.17	29.06	47.68	54.60	56.87	47.63
25	67.06	67.06	67.06	67.06	67.06	67.06	67.06
26	63.94	63.94	64.16	65.17	55.34	65.93	63.08
27	63.13	63.54	63.19	63.19	63.07	62.69	63.14
28	63.82	61.14	63.50	67.06	62.63	59.65	62.97
29	57.36	58.57	58.01	58.02	59.11	59.69	58.46
30	63.15	56.67	57.16	61.82	56.47	67.06	60.39
31	60.75	47.62	41.61	63.80	60.42	61.17	55.90
32	58.36	58.35	56.98	55.05	54.94	54.39	56.35
35	59.93	62.40	61.46	38.94	51.08	67.06	56.81

Key
Traditional Method
Game Application

Baselines	
Low	0
High	67.06

a. Two-Sample t-Test

We wanted to perform a two-sample t-test on the data but we did not meet the assumptions to do so.

Two Sample t-Test Assumptions:

(1) Is the data continuous or ordinal?

As seen in Figure 29 of Appendix C. Experiment Supplementals, the subjects annotated their confidence level in a continuous fashion by marking a single vertical line along a bracketed horizontal line representing “Not Confident” and “Extremely Confident” at the end stops.

(2) Are the subjects randomly selected?

The subjects were randomly placed in alternate groups based on arrival to the briefing venue as well as when each subject completed their demographic survey.

(3) Does the data represent a normal distribution?

Displayed in Figure 17, the histogram and quantile plots for the subjects’ overall average confidence levels displays that much of the data does not follow a normal distribution. Referencing the quantile plot, the majority of points do not fall along the expected eccentricity (the diagonal) of a normal distribution and there exists a left skew in the traditional condition’s histogram.

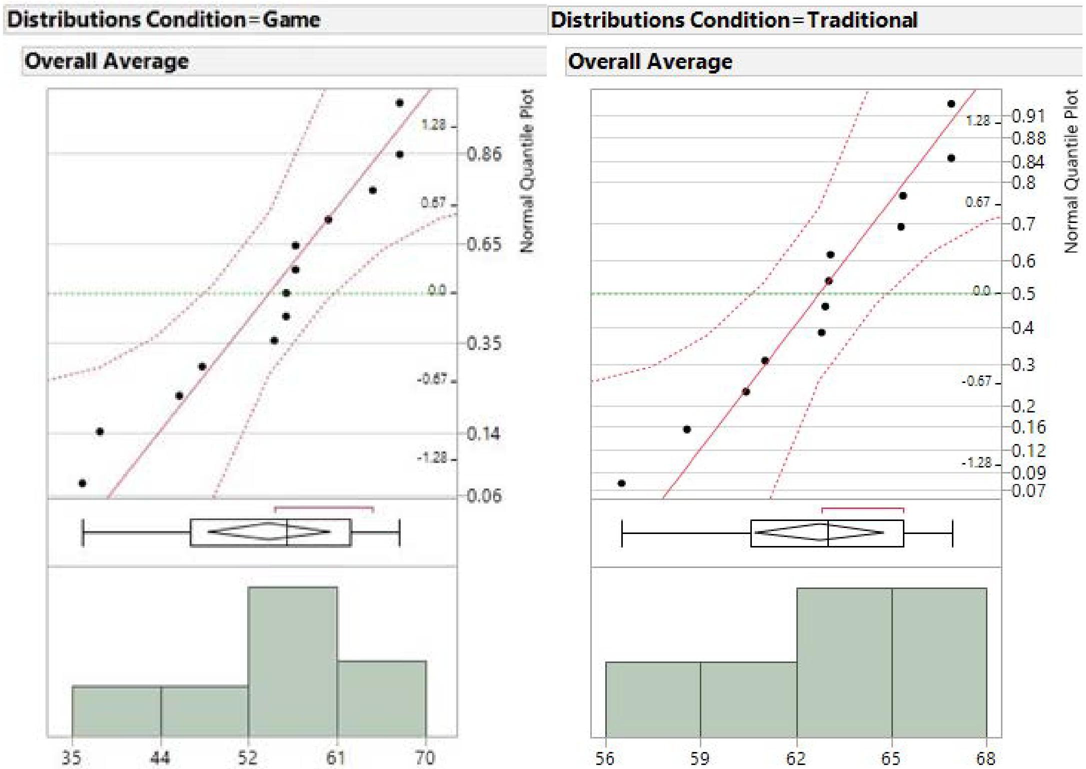


Figure 17. Histograms, Box Plots, and Quantile Plots for Subjects' Overall Average Confidence Levels

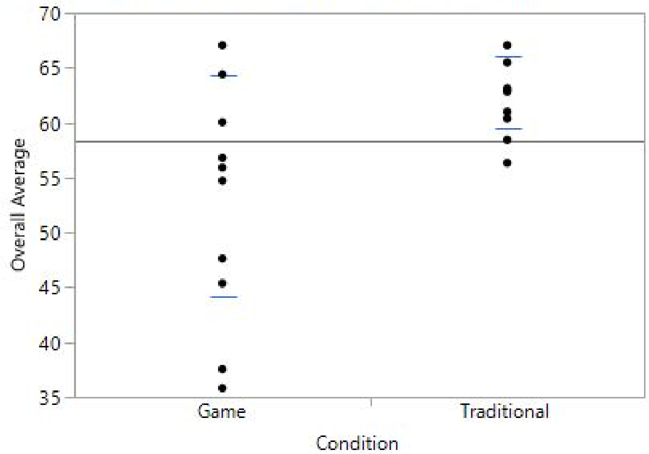
(4) Is there a reasonably large sample size?

Though there were 25 participants yielding 13 subjects for the game condition and 12 for the traditional condition, we can assume the sample size was not large enough since the distribution of results did not approach normality.

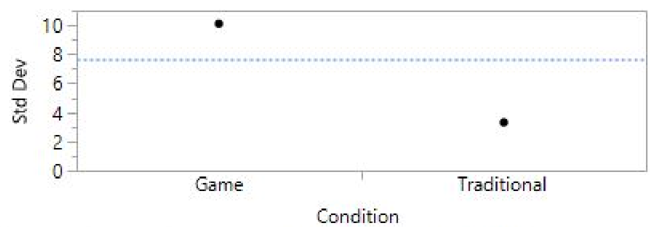
(5) Is there homogeneity of variance?

Homogeneity of variance will exist when the standard deviation of the two methods are approximately equal. The summary statistics shown in Figure 18 display a standard deviation of 10.1 for the game condition and 3.3 for the traditional condition. The difference in the standard deviations as well as the p-value for the Brown-Forsythe test leads to the conclusion that the variances are in fact different.

Oneway Analysis of Overall Average By Condition



Tests that the Variances are Equal



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
Game	13	10.10093	7.780079	7.322179
Traditional	12	3.30562	2.481944	2.436806

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	6.0828	1	23	0.0215*
Brown-Forsythe	5.5851	1	23	0.0269*
Levene	8.3282	1	23	0.0083*
Bartlett	11.2731	1	.	0.0008*
F Test 2-sided	9.3372	12	11	0.0008*

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

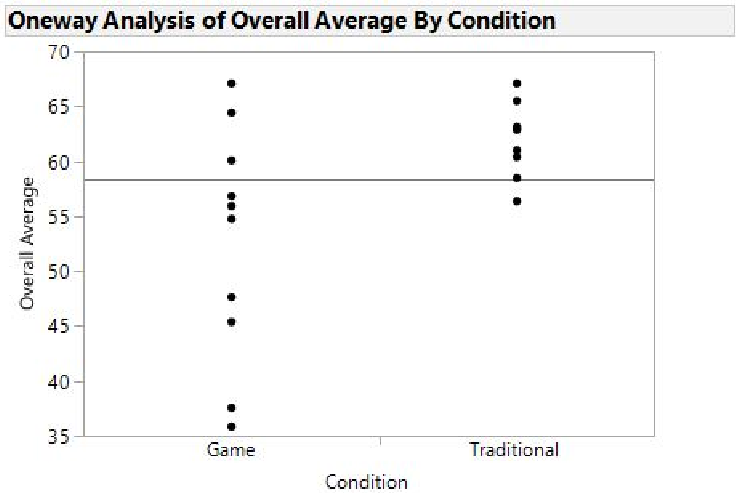
F Ratio	DFNum	DFDen	Prob > F
8.3076	1	14.73	0.0116*

t Test
2.8823

Figure 18. Analysis of Variance for Subjects' Overall Average Confidence Levels

b. *Wilcoxon Ranked Sum Method*

Since we fail to meet the assumptions and conditions in the areas of normality, sample size and homogeneity of variances for both the equal and unequal variance t-Test, we have to use the non-parametric Wilcoxon test, and the results are shown in Figure 19.



Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
Game	13	125.000	169.000	9.6154	-2.371
Traditional	12	200.000	156.000	16.6667	2.371

2-Sample Test, Normal Approximation

S	Z	Prob>|Z|
200	2.37065	0.0178*

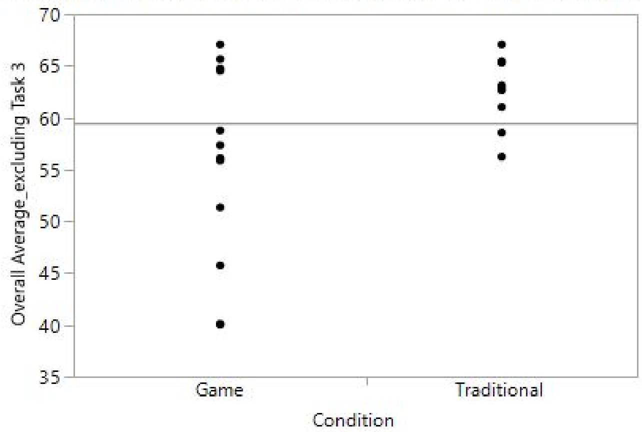
1-Way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
5.7499	1	0.0165*

Figure 19. Wilcoxon Rank Sum Analysis for Subjects' Overall Average Confidence Levels

After conducting the Wilcoxon test on each individual task, with results shown in Figure 38 through Figure 43 of Appendix F. Post-Task Confidence Analysis Results, reviewing the data and the SME evaluation comments, we noted task 3 had a potential discrepancy, which we cover in the discussion section. We conducted the analysis again for the subjects' overall average with task 3 removed; results are shown in Figure 20.

Oneway Analysis of Overall Average_excluding Task 3 By Condition



Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
Game	13	142.000	169.000	10.9231	-1.444
Traditional	12	183.000	156.000	15.2500	1.444

2-Sample Test, Normal Approximation

S	Z	Prob>|Z|
183	1.44419	0.1487

1-Way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
2.1651	1	0.1412

Figure 20. Wilcoxon Rank Sum Analysis for Subjects' Overall Average Confidence Levels (Excluding Task 3)

C. DISCUSSION

Discussion of the results provided in the previous section will be centered on determining if the developed application can be deemed capable of replacing or supplementing the current training method by answering the following questions:

- Is the application reliable?
- Are the subjects confident in the use of this application?
- What is the credibility of the usability study?

1. Results of Survey Data

Eleven total subjects, ranging from the ranks of E-3 to E-8, 3 to 28 years of service, and 21 to 47 years of age, participated in the usability survey. These subjects were all either current plane captains or had previously qualified as a plane captain and indicated that they were confident in their ability to perform the functions required of the job.

Each of the thirty survey items was posed in a format where a positive response is associated to a larger number on the provided scale, with scale values ranging from 1 to 5. Referencing Table 8, found in Appendix D. Usability Study Analysis Results, it can be noted that only 6 of 330 (1.8%) total responses fell below the midpoint value of 3, which indicates an overall positive reaction to the developed application. No survey item had an average response below 3.7 and no construct had an average response below 4.15. These values are in favor of the null hypothesis (H_0: $\mu C(x) \geq 3.00$), supporting the premise that participants find the developed application useful, easy to use, easy to learn, and satisfying.

2. Reliability of Survey Data

After examining the survey results, we wanted to check that our data was reliable as well as non-redundant. We performed Cronbach’s Alpha test, which indicated we should exclude three of our survey items:

- Item 3: Is it useful?
- Item 8: It does everything I would expect it to do.
- Item 27: It works the way I want it to work.

With these items removed, the results from these subjects’ surveys presented reliability coefficients above 0.70 in all constructs.

The comments provided by the usability survey participants provide insight into the problematic nature of these questions. These problems can be tied to comments relating to redundancy, the prototype nature of the application, and to the area of feedback.

Item three can be seen as redundant as the surrounding questions of that set note similar items but more explicit in specified area of usefulness to the participant. For item 8, user comments indicated that the subjects believe that it should provide much more functionality than the current prototype provides. Additionally, users commented on problems due to the prototype nature of the system: the system’s inability to recognize specified movements, inaccuracies in counting the number of correct movements, and glitches found in certain movements. Though item 27 also correlates to the system's prototypical nature, this item can also be tied specifically to the application's inability to provide the appropriate feedback. Comments of this nature included statements on the application's inability to provide remediation on errors as well details missing from the provided task.

The final reliability coefficients, shown in Table 5 display empirical evidence that the responses to the scale are in favor of the null hypothesis, meaning they are reliable, consistent, and thus credible. This analysis continues to support the construct categories in that the application of a virtualized training system for this qualification, and potentially other aviation maintenance tasks, will be useful, easy to interact with, easy to learn, and satisfying.

3. Kinect Reliability

The initial split of the data within the partition tree occurs between the Kinect and the SME pass rates; this split supports the rejection of the null hypothesis (H_0: $P_K = P_{SME}$) in favor of the alternative hypothesis (H_a: $P_K \neq P_{SME}$) that the average pass rates for the Kinect and the SMEs are not equal. Looking at the results, the SMEs graded more motions as correct than the Kinect, indicating that the Kinect had significant false negatives.

The partition tree also identifies splits between the game and traditional conditions as graded by both SMEs and the Kinect. This suggests that there is also a difference between the pass rates of the users based on the medium in which the material was taught, indicating that the experimental condition did not train as well as the traditional methods.

Looking at splits #3, #4, #6, and #7, the Kinect and the SMEs do concur on specified tasks in the lower bracket of the game condition and the higher bracket of the traditional condition.

The concurrences in the lower bracket of the game condition are task 1, task 3, and task 4 which are corroborated by comments provided by the SMEs as well as the specified parameters programed into the application. The common themes of the comments provided by all three SMEs for these tasks are centered on the subjects' hand placement, hand orientation, and the disposition of when the subjects' hands should be open or closed. These themes are thought to be found in the lower bracket of the game condition due to the beta nature of the application and comments identifying the need for more explicit direction in the application's instructional series. Nevertheless, these themes do fall in line with the significant elements that were programmed into the gesture recognition portion of the application for these tasks and demonstrate that the application is capable of identifying the same deficiencies as the SMEs.

On the higher end of the traditional condition, concurrences are found for task 2, task 3, and task 6. In order to address why these concurrences are found on the high end of the traditional condition the comments provided on the opposite spectrum for these tasks, relating to conduct in the game condition, are reviewed. These comments once again identify deficiencies found in in the subjects' hand placement and hand orientation but with the added deficiency in the subject's head orientation. Assuming that these deficiencies are addressed by the traditional method and these significant elements are appropriately programmed into the gesture recognition portion of the application, we have now demonstrated the application's ability to identify positive attributes of the subjects' ability to perform the specified task.

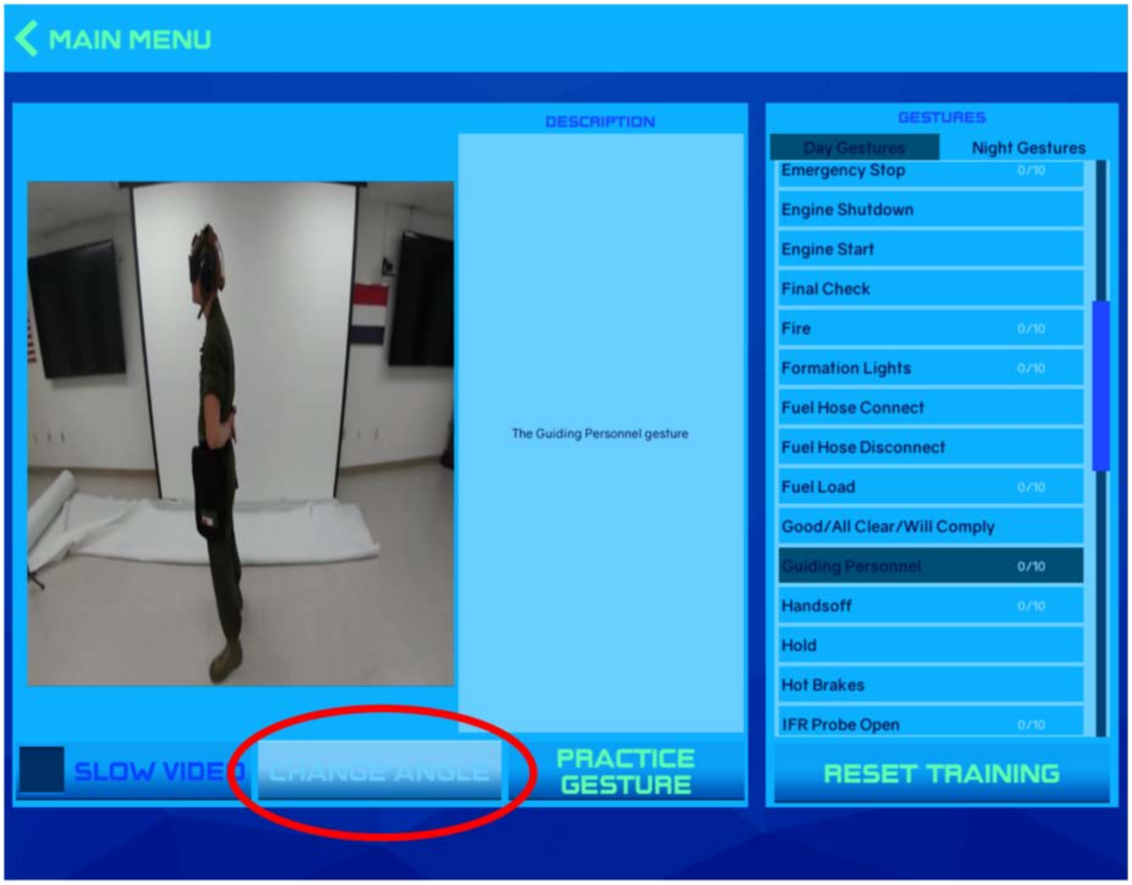
Considering the aforementioned factors, it can be presumed that there is the capability for this application or similar applications to perform as reliably as the SMEs in the ability to assess both adequate and deficient task completions. These capabilities have the potential for standardization in both the realm of instruction and computer-based assessment.

4. Trainee Confidence

Examining the difference in the mean value of subjects' confidence to perform each of the measured gestures yielded interesting results. The p-value from the 1-way Test, Chi Square Approximation shown in Figure 19 is 0.0165, which is lower than our chosen alpha of 0.05. Thus, we reject the null hypothesis, $\mu_T = \mu_G$, in favor of the alternative hypothesis, $\mu_T \neq \mu_G$. Taking this p-value and the distribution of confidence markings into account it can be concluded that the subjects were overall less confident utilizing the application to learn the selected movements.

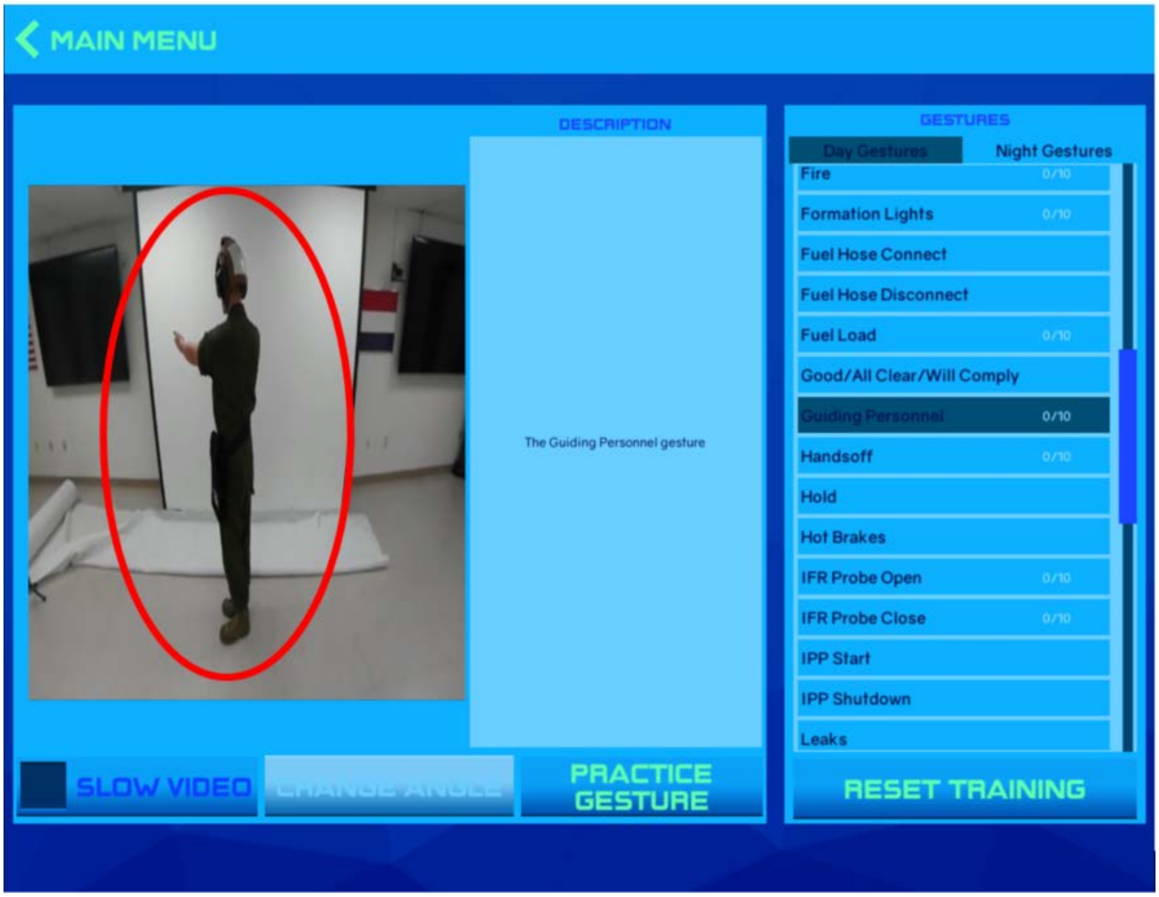
However, one data point appeared suspect. The p-value from the 1-way Test, Chi Square Approximation with task 3 removed shown in Figure 20 is 0.1412. The value of 0.1412 is greater than a value of 0.05, thus failing to reject the null hypothesis ($\mu_T = \mu_G$). Taking this p-value and the distribution of confidence markings with task 3 removed into account, it can be concluded that the subjects' average confidence levels utilizing the application to learn the selected movements were equal to that of the traditional method with the exception of task 3.

Using the analysis of the subject' overall confidence levels with task 3 removed, reviewing the material provided, and considering the comments provided by the SMEs when grading this particular movement, it can be inferred that the participants were less confident with this task due to discrepancies in the provided lesson. This particular lesson displayed a lack of feedback by the application, as well as lack of guidance provided to the SMEs. The average confidence level for task 3 for those utilizing the game application ranged from 11 to 16 units lower than the other tasks. After reviewing the material provided by the application, it can be noted that only one view (the side view) of this lesson was available, shown in Figure 21, while the other lessons were afforded two perspectives, one from the front and one from the side profiles. The SMEs also commented on failures due to subjects turning at the torso, which was deemed acceptable by the provided material shown in Figure 22. These discrepancies merit consideration for conducting the second analysis of confidence levels and concluding that users are generally confident utilizing the application.



Note: This particular task only displayed one perspective due to loss of data.

Figure 21. Screenshot of Task 3 Instructional Series



Note: Video instruction accommodated turning at the torso.

Figure 22. Screenshot of Task 3 Instructional Series

VI. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSION

The following research questions were the focal point for this thesis:

1. Is it feasible to create and utilize a game-based part-task trainer to replace or supplement the current method for training aircraft plane captains? We break this down into three sub-questions: 1) can a trainer built using commercial off the shelf (COTS) technology determine whether a trainee performs hand and arm signals correctly as well as subject-matter experts (SMEs)?; 2) can such a trainer produce trained plane captains as well as the traditional methods of instruction?; 3) Is there a difference between the confidence of plane captains trained in the different methods?

2. Will users consider a virtualized game-based learning platform a practical means of increasing sets and reps to attain the qualification of plane captain? We answer this question by surveying experienced plane captains who have used the system.

Unfortunately, the results of this thesis are such that we cannot answer any of question 1's three sub-questions with a "Yes." The Kinect trainer produced false negatives compared to a group of SMEs when evaluating trainees (1.1), the plane captains trained in the traditional methods outperformed those trained with the developed application (1.2), and there was no statistically significant difference between the confidence of the two groups (1.3).

However, the fact that question 2 was answered with a very significant "Yes" by experienced plane captains and some of the data lead us to believe that the reason for the failure of the virtualized game-based training to succeed was essentially due to time constraints and the prototypical nature of the system. In comparison to an actual product development, what we produced was essentially the first cut of a final product. Had this been an actual product, after these initial results, the system would have been further engineered to increase the reliability of the evaluation of user's gestures.

The task analysis, system design, and prototype development demonstrated the feasibility of creating a virtualized plane captain training system. Utilizing a combination of a COTS game engine, 360 video capture, and the Microsoft Kinect, we were able to provide the user with an immersive environment that not only presented the material, but also afforded the user the ability to practice and visualize the effects of their actions.

Referring back to the questions utilized for discussion, we can outline elements of the developed application that denote credibility, reliability, and how confident the users were in the developed application. Utilizing an internal consistency calculation, we were able to determine that the application of a virtualized training system is indeed useful, easy to interact with, easy to learn, and satisfying. Though this particular implementation of a virtualized trainer was not determined to be reliable due to its prototypical nature, with improved hardware and software, the capability to perform with increased reliability is easily achievable. Correlating explicit deficiencies to a lack in confidence in a specified task, the users of this system were otherwise confident in the execution of the developed application and theory supporting it.

Overall, we have shown that it is certainly possible to create a virtualized plane captain training system. With improvements to the execution, this or similar systems can be a viable means to replace or supplement the current training methodologies.

B. FUTURE WORK

Due to the prototypical nature of the development of this application and the potential impact that the underlying theory would have on current training methodologies, there are many opportunities to improve or expand on this research. The following are suggested improvements to the developed system and opportunities for expanded research:

1. Suggested Improvements to the Developed System

- Utilize higher fidelity tracking hardware for improved feedback. This feedback would potentially include specific identification of user deficiencies.

- Enhance the immersion capability through additional display options, such as a head mounted display.

- Add additional scenarios to the procedural series for advanced use. Suggestions include the incorporation of emergency scenarios.

- Display instructional notes to clarify important aspects within the instructional series. Examples may include specifying a dominant arm or denoting where the user should look. These notes would clarify items not readily apparent in the video instruction.

2. Opportunities for Expanded Research

- Implement the underlying theory to other areas of training, to include disciplines and services outside of Marine aviation maintenance.

- Incorporate performance based scoring to illicit competition and conduct research on the impact to user performance and retention.

- Conduct prolonged research to assess transfer of training.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. TASK ANALYSIS RESULTS

Task 1:

Assumptions:

(1) Individuals have successfully passed basic military training.
(2) Individuals have completed their primary military occupational specialty school as an aviation maintenance technician.
(3) This task is being completed as a part of normal garrison squadron operations, for pilot training.
(4) Post/before operations servicing inspections have been completed the night before.

Limitations:

(1) This task does not include ancillary checks for persistent maintenance issues attributed to specific aircraft.

Goal 1. Start Engine
Method for Goal 1.0
Goal 1.1 Prepare aircraft
Method for Goal 1.1
Operator: Identify and carry appropriate tools (motor)
Operator: Identify and don appropriate personal protective equipment PPE (motor)
Operator: Walk out to identified aircraft (motor)
Operator: Visually inspect aircraft for abnormalities (perceptual)
Operator: Stow red gear pins/covers/plugs (motor)
Operator: Conduct debrief with pilot (cognitive)
Operator: Close ladder bay door (motor)

Goal 1.2 Position/Posture for engine start up
Method for Goal 1.2
Operator: Walk 15ft to the left of the nose of the aircraft (motor)
Operator: Walk 2ft forward of the nose of the aircraft (motor)
Operator: Turn to face aircraft and pilot (motor)
Operator: Assume parade rest (motor)
Operator: Confirm pilot signal integrated power pack start (perceptual)
Operator: Confirm integrated power package exhaust is clear (perceptual/cognitive)
Operator: Reciprocate pilot signal (motor)
Operator: Assume parade rest (motor)
Operator: Confirm integrated power package start (perceptual)

Figure 23. Task Analysis for the Start Engine Hand-and-Arm Signal

Goal 1.3 Close Canopy
Method for Goal 1.3
    Operator: Confirm pilot signal canopy close (perceptual)
    Operator: Confirm canopy railing is clear (perceptual/cognitive)
    Operator: Reciprocate pilot signal (motor)
    Operator: Assume parade rest (motor)

Goal 1.4 Prepare engine start
Method for Goal 1.4
    Operator: Recognize pilot signal for engine start (cognitive/perceptual)
    Operator: Stand at attention (motor)
    Operator: Make a fist with right hand (motor)
    Operator: Raise right hand to head level perpendicular to ground (motor)
    Operator: Maintain for duration of safety checks (motor)

Goal 1.5 Conduct safety checks
Method for Goal 1.5
    Operator: Confirm personnel safety distance (cognitive)
    Operator: Confirm equipment safety distance (cognitive)
    Operator: Confirm aft of exhaust clear (cognitive)

Goal 1.6 Give signal for engine start up
Method for Goal 1.6
    Operator: Raise left hand overhead at 45 degree angle (motor)
    Operator: Extend finger of left hand to indicate desired engine number to start (motor/cognitive)
    Operator: Create fist with right hand (motor)
    Operator: Raise right fist and elbow perpendicular to shoulder (motor)
    Operator: Maintain elbow shoulder alignment (motor)
    Operator: Rotate fist counterclockwise, pivoting at the elbow (motor)

Goal 1.7 Pilot starts engine
Method for Goal 1.7
    Operator: Confirm begin of engine start (perceptual)
    Operator: Assume parade rest (motor)

Figure 24. Task Analysis for the Start Engine Hand-and-Arm Signal (Continued)

Task 2:

Assumptions:

(1) Individuals have successfully passed basic military training.
(2) Individuals have completed their primary military occupational specialty school as an aviation maintenance technician.
(3) This task is being completed as a part of normal garrison squadron operations, for pilot training.
(4) Post/before operations servicing inspections have been completed the night before.
(5) Engine start procedures have been accomplished
(6) PC is positioned at parade rest at the appropriate location

Limitations:

(1) This task does not include ancillary checks for persistent maintenance issues attributed to specific aircraft.

Goal 2. Signal engine fire
Method for Goal 2.0
- Goal 2.1 Confirm engine fire (perceptual/cognitive)
- Method for Goal 2.1
  - Operator: Identify abnormal smoke accumulation (perceptual) or
  - Operator: Identify flames from the exhaust (perceptual)
- Goal 2.2 Signal engine fire
- Method for Goal 2.2
  - Goal 2.2.1 Signal location of engine fire
  - Method for Goal 2.2.1
    - Operator: Raise left hand overhead at 45 degree angle (motor)
    - Operator: Extend finger of left hand to indicate location of fire (motor/cognitive)
    - Operator: Maintain left hand position for duration of movements
  - Goal 2.2.2 Motion figure 8 with right hand
  - Method for Goal 2.2.2
    - Operator: Extend right arm forward above right shoulder at head level (motor)
    - Operator: Move right arm across the body extended downward past the left hip (motor)
    - Operator: Move right arm upward above the left shoulder at head level (motor)
    - Operator: Move right arm across the body extended downward past the right hip (motor)
    - Operator: Operator: Move right arm upward above the right shoulder at head level (motor)
    - Operator: Continue methods for Goal 2.2.2 until pilot acknowledgement (motor/cognitive)

Figure 25. Task Analysis for the Engine Fire Hand-and-Arm Signal

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. SYSTEM DESIGN SUPPLEMENTALS

Table 7. System Design Constraint Swing Weight

Swing Weights					
	Consequence to compare				
Attribute swung from worst to best	**Cost**	**Mod**	**Ease**	**Throughput**	**Network**
Benchmark	5	3	7	8	8
Cost	10	3	7	7	8
Modular	5	10	7	7	8
Ease of use	5	3	10	7	8
Throughput	5	3	7	10	8
Network Access	5	3	7	7	10
Interoperability	5	3	7	8	8
Size	5	3	7	9	8
Lifecycle Maintainability	5	3	7	10	8

Swing Weights					
Consequence to compare			**Rank**	**Rate**	**Weight**
Interoperaility	**Size**	**Life**			
4	8	6	9	0	0
4	8	6	1	100	16.39%
4	8	6	3	90	14.75%
4	8	6	2	90	14.75%
4	8	6	4	80	13.11%
4	8	6	7	60	9.84%
10	8	6	5	60	9.84%
4	10	6	6	70	11.48%
4	8	10	8	60	9.84%
			Sum =	610	100%

Note: Swing weights are determined utilizing the Clemen and Reilly method

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. EXPERIMENT SUPPLEMENTALS

Plane Captain Training System

Demographic Survey

Subject Number: _______ Date: _______

1. Rank: ____________
2. Years of Service: ____________
3. Age: ____________
4. Gender: Female Male
5. What is your primary Military Occupational Specialty (MOS):
   a. MOS designation number (i.e. 60XX): ____________
   b. MOS designation title (i.e. Joint Strike Fighter (JSF) Maintenance Specialist):

      ____
6. Rate your ability to perform the Plane Captain duties.
   Circle one:

   1 2 3 4 5

   Little to no knowledge … Able to launch an aircraft flawlessly
7. Select the number of Plane Captain Hand & Arm Signals that you are able to perform from memory.
   Circle one:

   0 1 – 5 6 – 10 11 – 15 16 +
8. Do you play video games? Yes No

   If Yes…

   a. How often? < 2 hrs/wk 2 – 4 hrs/wk 4 – 8 hrs/wk > 8 hrs/wk
   b. What Kinds? single player multi player first-person third-person

      Other: ________________
   c. Have you used gesture recognition games, such as:

      XBOX Kinect, PlayStation Move, Wii, or other similar system ______________
      (Please annotate systems not listed)

      If yes, estimate total time played?

      < 2 hrs 2 – 10 hrs 11 – 20 hrs > 20 hrs

Figure 26. Participant Demographic Survey

Plane Captain Training System Usability Study

Based on: Lund, A.M. (2001) *Measuring Usability with the USE Questionnaire.* STC Usability SIG Newsletter, 8:2. [Abstract] | About quest.cgi

Please rate your agreement with these statements.

- Try to respond to all the items.
- For items that are not applicable, use: **N/A**
- Make sure these fields are filled in: **Webform:** **Email to:**
- Add a comment about an item by clicking on its icon, or add comment fields for all items by clicking on **Comment All**.
- To mail in your results, click on: **Mail Data**

Webform: Subject ID# **Email to:** ellandas@nps.edu

Optionally provide comments and your email address in the box.

[Mail Data] [Comment All]

		-2	-1	0	1	2		N/A
USEFULNESS		-2	-1	0	1	2		N/A
1. It helps me be more effective.	DISAGREE	○	○	○	○	○	AGREE	○
2. It helps me be more productive.	DISAGREE	○	○	○	○	○	AGREE	○
3. It is useful.	DISAGREE	○	○	○	○	○	AGREE	○
4. It gives me more control over the activities in my life.	DISAGREE	○	○	○	○	○	AGREE	○
5. It makes the things I want to accomplish easier to get done.	DISAGREE	○	○	○	○	○	AGREE	○
6. It saves me time when I use it.	DISAGREE	○	○	○	○	○	AGREE	○
7. It meets my needs.	DISAGREE	○	○	○	○	○	AGREE	○
8. It does everything I would expect it to do.	DISAGREE	○	○	○	○	○	AGREE	○
EASE OF USE		-2	-1	0	1	2		N/A
9. It is easy to use.	DISAGREE	○	○	○	○	○	AGREE	○
10. It is simple to use.	DISAGREE	○	○	○	○	○	AGREE	○
11. It is user friendly.	DISAGREE	○	○	○	○	○	AGREE	○
12. It requires the fewest steps possible to accomplish what I want to do with it.	DISAGREE	○	○	○	○	○	AGREE	○
13. It is flexible.	DISAGREE	○	○	○	○	○	AGREE	○
14. Using it is effortless.	DISAGREE	○	○	○	○	○	AGREE	○
15. I can use it without written instructions.	DISAGREE	○	○	○	○	○	AGREE	○
16. I don't notice any inconsistencies as I use it.	DISAGREE	○	○	○	○	○	AGREE	○
17. Both occasional and regular users would like it.	DISAGREE	○	○	○	○	○	AGREE	○
18. I can recover from mistakes quickly and easily.	DISAGREE	○	○	○	○	○	AGREE	○
19. I can use it successfully every time.	DISAGREE	○	○	○	○	○	AGREE	○
EASE OF LEARNING		-2	-1	0	1	2		N/A
20. I learned to use it quickly.	DISAGREE	○	○	○	○	○	AGREE	○
21. I easily remember how to use it.	DISAGREE	○	○	○	○	○	AGREE	○
22. It is easy to learn to use it.	DISAGREE	○	○	○	○	○	AGREE	○
23. I quickly became skillful with it.	DISAGREE	○	○	○	○	○	AGREE	○
SATISFACTION		-2	-1	0	1	2		N/A
24. I am satisfied with it.	DISAGREE	○	○	○	○	○	AGREE	○
25. I would recommend it to a friend.	DISAGREE	○	○	○	○	○	AGREE	○
26. It is fun to use.	DISAGREE	○	○	○	○	○	AGREE	○
27. It works the way I want it to work.	DISAGREE	○	○	○	○	○	AGREE	○
28. It is wonderful.	DISAGREE	○	○	○	○	○	AGREE	○
29. I feel I need to have it.	DISAGREE	○	○	○	○	○	AGREE	○
30. It is pleasant to use.	DISAGREE	○	○	○	○	○	AGREE	○
		-2	-1	0	1	2		N/A

List the most **negative** aspect(s):

1.
2.
3.
4.
5.

List the most **positive** aspect(s):

1.
2.
3.
4.
5.

Figure 27. Plane Captain Training System Usability Study Survey

Plane Captain Training System

Subject Matter Expert (SME) Scoring Sheet

Subject Number: _______ Rater ID # _______ Date: _______

Movement

1. ____________________

rep 1:	PASS	FAIL	rep 6:	PASS	FAIL
rep 2:	PASS	FAIL	rep 7:	PASS	FAIL
rep 3:	PASS	FAIL	rep 8:	PASS	FAIL
rep 4:	PASS	FAIL	rep 9:	PASS	FAIL
rep 5:	PASS	FAIL	rep 10:	PASS	FAIL

2. ____________________

rep 1:	PASS	FAIL	rep 6:	PASS	FAIL
rep 2:	PASS	FAIL	rep 7:	PASS	FAIL
rep 3:	PASS	FAIL	rep 8:	PASS	FAIL
rep 4:	PASS	FAIL	rep 9:	PASS	FAIL
rep 5:	PASS	FAIL	rep 10:	PASS	FAIL

3. ____________________

rep 1:	PASS	FAIL	rep 6:	PASS	FAIL
rep 2:	PASS	FAIL	rep 7:	PASS	FAIL
rep 3:	PASS	FAIL	rep 8:	PASS	FAIL
rep 4:	PASS	FAIL	rep 9:	PASS	FAIL
rep 5:	PASS	FAIL	rep 10:	PASS	FAIL

4. ____________________

rep 1:	PASS	FAIL	rep 6:	PASS	FAIL
rep 2:	PASS	FAIL	rep 7:	PASS	FAIL
rep 3:	PASS	FAIL	rep 8:	PASS	FAIL
rep 4:	PASS	FAIL	rep 9:	PASS	FAIL
rep 5:	PASS	FAIL	rep 10:	PASS	FAIL

5. ____________________

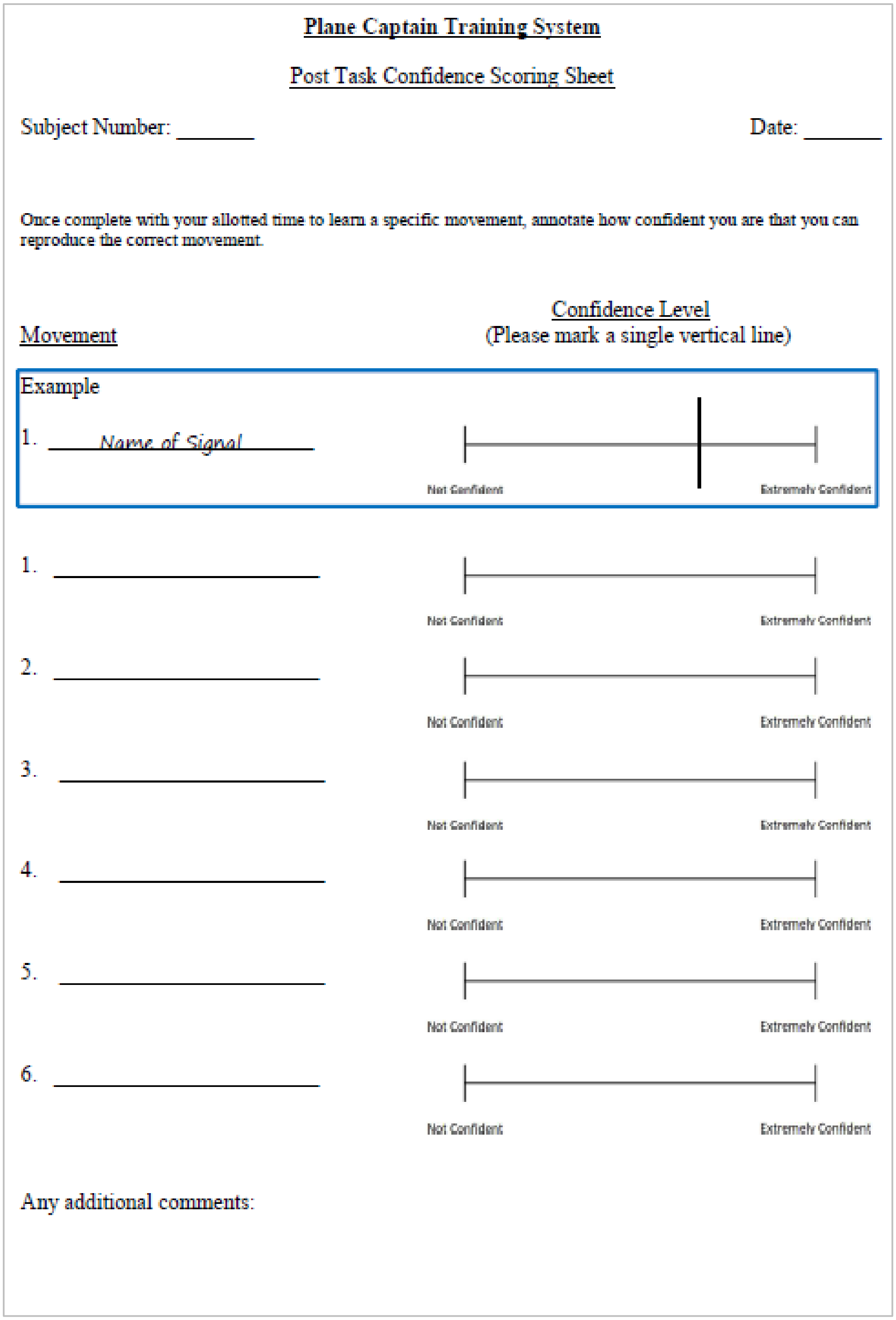
rep 1:	PASS	FAIL	rep 6:	PASS	FAIL
rep 2:	PASS	FAIL	rep 7:	PASS	FAIL
rep 3:	PASS	FAIL	rep 8:	PASS	FAIL
rep 4:	PASS	FAIL	rep 9:	PASS	FAIL
rep 5:	PASS	FAIL	rep 10:	PASS	FAIL

6. ____________________

rep 1:	PASS	FAIL	rep 6:	PASS	FAIL
rep 2:	PASS	FAIL	rep 7:	PASS	FAIL
rep 3:	PASS	FAIL	rep 8:	PASS	FAIL
rep 4:	PASS	FAIL	rep 9:	PASS	FAIL
rep 5:	PASS	FAIL	rep 10:	PASS	FAIL

Any additional comments:

Figure 28. SME Scoring Sheet


Plane Captain Training System

Post Task Confidence Scoring Sheet

Subject Number: _______ Date: _______

Once complete with your allotted time to learn a specific movement, annotate how confident you are that you can reproduce the correct movement.

Movement

Confidence Level
(Please mark a single vertical line)

Example

1. Name of Signal

Not Confident Extremely Confident

1. ____________ Not Confident Extremely Confident

2. ____________ Not Confident Extremely Confident

3. ____________ Not Confident Extremely Confident

4. ____________ Not Confident Extremely Confident

5. ____________ Not Confident Extremely Confident

6. ____________ Not Confident Extremely Confident

Any additional comments:

Figure 29. Post-task Confidence Scoring Sheet

APPENDIX D. USABILITY STUDY ANALYSIS RESULTS

Table 8. Usability Study Results

Subject ID #	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
12	4	5	4	3	4	5	4	5
14	4	4	4	5	4	5	NA	4
15	3	4	5	3	4	3	2	5
16	4	4	5	5	4	4	5	4
17	4	5	5	3	5	5	5	4
18	4	4	4	4	4	4	4	4
19	4	4	5	3	4	4	3	4
36	5	5	5	5	5	5	5	4
37	4	4	5	3	4	4	3	4
38	5	4	5	5	5	4	5	5
39	5	5	4	4	5	3	3	2

Key
Usefulness
Ease of Use
Ease of Learning
Satisfaction

Subject ID #	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19
12	5	5	5	5	4	5	5	4	4	3	4
14	4	5	5	5	5	4	3	4	4	4	NA
15	5	5	5	4	3	3	1	2	4	5	3
16	4	4	5	4	4	4	5	4	4	4	4
17	5	5	5	5	4	5	5	4	5	4	5
18	4	4	4	4	4	4	4	4	4	4	4
19	4	4	4	3	4	3	4	4	4	4	4
36	5	5	4	4	5	4	4	4	5	4	5
37	4	4	4	3	4	3	4	4	4	4	4
38	5	5	5	5	5	5	5	2	4	5	5
39	4	4	5	4	3	4	5	5	5	5	5

Subject ID #	Q20	Q21	Q22	Q23
12	5	5	5	4
14	4	4	5	5
15	4	4	4	4
16	5	5	5	5
17	5	5	5	5
18	4	4	4	4
19	4	4	4	5
36	5	5	5	5
37	4	4	4	5
38	5	5	5	5
39	4	5	5	5

Subject ID #	Q24	Q25	Q26	Q27	Q28	Q29	Q30
12	5	5	5	4	5	5	5
14	5	5	4	4	4	3	4
15	2	4	5	5	3	4	5
16	3	4	4	4	3	3	4
17	5	5	5	4	5	3	5
18	4	4	4	4	4	4	4
19	4	4	5	4	4	4	4
36	3	5	5	3	5	NA	5
37	4	4	5	4	4	4	4
38	5	5	5	4	4	4	5
39	3	5	5	2	3	3	3

Note: Subject marking have been translated from a -2 to 2 scale to a 1 to 5 scale for analysis purposes

Table 9. Usability Study Total Average User Response

Underlying Construct	Item Number								Overall Average
Usefulness	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	μC(Usefulness)
Overall Average User Response	4.18	4.36	4.64	3.91	4.36	4.18	3.90	4.09	4.20

Underlying Construct	Item Number											Overall Average
Ease of Use	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	μC(Ease of Use)
Overall Average User Response	4.45	4.55	4.64	4.18	4.09	4.00	4.09	3.73	4.27	4.18	4.30	4.23

Underlying Construct	Item Number				Overall Average
Ease of Learning	Q20	Q21	Q22	Q23	μC(Ease of Learning)
Overall Average User Response	4.45	4.55	4.64	4.73	4.59

Underlying Construct	Item Number							Overall Average
Satisfaction	Q24	Q25	Q26	Q27	Q28	Q29	Q30	μC(Satisfaction)
Overall Average User Response	3.91	4.55	4.73	3.82	4.00	3.70	4.36	4.15

Multivariate

Correlations

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Q1	1.0000	0.4183	-0.0896	0.5590	0.7470	0.1406	0.3462	-0.4353
Q2	0.4183	1.0000	-0.2143	-0.1336	0.6071	0.3360	0.0694	-0.3252
Q3	-0.0896	-0.2143	1.0000	-0.0764	0.1786	-0.0720	-0.1439	0.3252
Q4	0.5590	-0.1336	-0.0764	1.0000	0.2864	0.1668	0.6992	-0.1159
Q5	0.7470	0.6071	0.1786	0.2864	1.0000	0.0720	0.2211	-0.3252
Q6	0.1406	0.3360	-0.0720	0.1668	0.0720	1.0000	0.7358	0.2914
Q7	0.3462	0.0694	-0.1439	0.6992	0.2211	0.7358	1.0000	0.0796
Q8	-0.4353	-0.3252	0.3252	-0.1159	-0.3252	0.2914	0.0796	1.0000

There are 1 missing values. The correlations are estimated by REML method.

Cronbach's α

	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.6161	

Excluded Col	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q1	0.5625	
Q2	0.6235	
Q3	0.6499	
Q4	0.5243	
Q5	0.5771	
Q6	0.5075	
Q7	0.4058	
Q8	0.6887	

Cronbach's α, standardized

	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.5953	

Excluded Col	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q1	0.5040	
Q2	0.5908	
Q3	0.6524	
Q4	0.5349	
Q5	0.4930	
Q6	0.5057	
Q7	0.4676	
Q8	0.6779	

Note: Question 3 and 8’s α is larger than the set in both the normal and standardized.

Figure 30. Analysis of Cronbach’s Alpha for the Usefulness Construct

Multivariate

Correlations

	Q1	Q2	Q4	Q5	Q6	Q7
Q1	1.0000	0.4183	0.5590	0.7470	0.1406	0.4062
Q2	0.4183	1.0000	-0.1336	0.6071	0.3360	0.1282
Q4	0.5590	-0.1336	1.0000	0.2864	0.1668	0.6953
Q5	0.7470	0.6071	0.2864	1.0000	0.0720	0.2968
Q6	0.1406	0.3360	0.1668	0.0720	1.0000	0.7405
Q7	0.4062	0.1282	0.6953	0.2968	0.7405	1.0000

There are 1 missing values. The correlations are estimated by REML method.

Multivariate Simple Statistics

Column	N	DF	Mean	Std Dev	Sum	Minimum	Maximum
Q1	11	10.00	4.1818	0.6030	46.0000	3.0000	5.0000
Q2	11	10.00	4.3636	0.5045	48.0000	4.0000	5.0000
Q4	11	10.00	3.9091	0.9439	43.0000	3.0000	5.0000
Q5	11	10.00	4.3636	0.5045	48.0000	4.0000	5.0000
Q6	11	10.00	4.1818	0.7508	46.0000	3.0000	5.0000
Q7	10	9.00	4.0598	1.1760	39.0000	2.0000	5.0000

Cronbach's α

	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.7598	

Excluded Col	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q1	0.7069	
Q2	0.7709	
Q4	0.7210	
Q5	0.7368	
Q6	0.7304	
Q7	0.6439	

Cronbach's α, standardized

	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.7748	

Excluded Col	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q1	0.7013	
Q2	0.7773	
Q4	0.7612	
Q5	0.7254	
Q6	0.7700	
Q7	0.7017	

Note: Question 3 and 8 were not included in this analysis.

Figure 31. Analysis of Cronbach's Alpha for the Usefulness Construct with Redundant or Erroneous Items Removed

Multivariate

Correlations

	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19
Q9	1.0000	0.8333	0.3105	0.5333	0.1491	0.4944	-0.0713	-0.5581	0.2609	0.0289	0.1562
Q10	0.8333	1.0000	0.4485	0.7420	0.3975	0.4944	-0.2424	-0.5004	0.1491	-0.0289	0.1562
Q11	0.3105	0.4485	1.0000	0.7200	-0.1800	0.5118	0.0590	-0.2390	0.0386	0.2390	0.0638
Q12	0.5333	0.7420	0.7200	1.0000	0.3457	0.8598	0.1983	-0.2142	0.1296	-0.0803	0.3793
Q13	0.1491	0.3975	-0.1800	0.3457	1.0000	0.3685	0.2232	-0.1148	-0.0833	-0.2797	0.4939
Q14	0.4944	0.4944	0.5118	0.8598	0.3685	1.0000	0.6344	-0.0000	0.2764	-0.2141	0.6049
Q15	-0.0713	-0.2424	0.0590	0.1983	0.2232	0.6344	1.0000	0.4774	0.3028	-0.2963	0.7232
Q16	-0.5581	-0.5004	-0.2390	-0.2142	-0.1148	-0.0000	0.4774	1.0000	0.4303	-0.4500	0.3297
Q17	0.2609	0.1491	0.0386	0.1296	-0.0833	0.2764	0.3028	0.4303	1.0000	0.1614	0.7157
Q18	0.0289	-0.0289	0.2390	-0.0803	-0.2797	-0.2141	-0.2963	-0.4500	0.1614	1.0000	0.1041
Q19	0.1562	0.1562	0.0638	0.3793	0.4939	0.6049	0.7232	0.3297	0.7157	0.1041	1.0000

There are 1 missing values. The correlations are estimated by Pairwise method.

Cronbach's α

	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.7085	

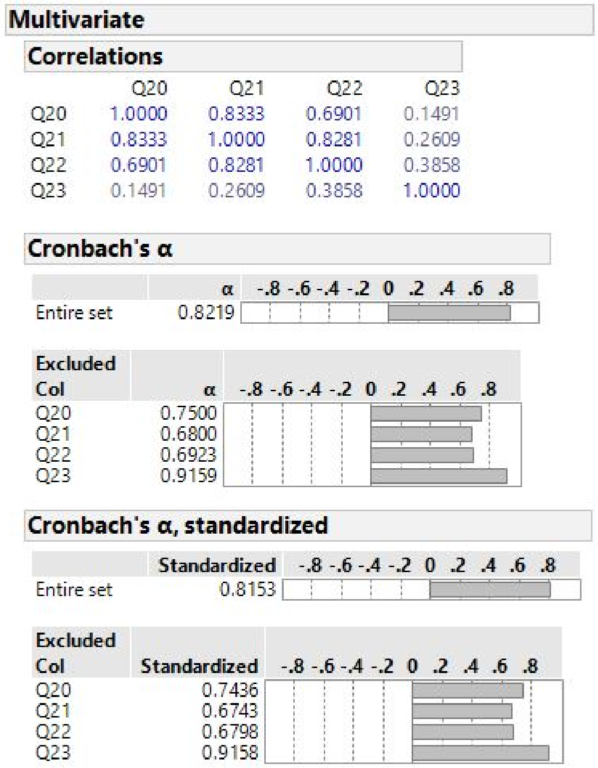
Excluded Col	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q9	0.6972	
Q10	0.6926	
Q11	0.6937	
Q12	0.6399	
Q13	0.7002	
Q14	0.5981	
Q15	0.6770	
Q16	0.7608	
Q17	0.6795	
Q18	0.7567	
Q19	0.6147	

Cronbach's α, standardized

	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.7332	

Excluded Col	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q9	0.7101	
Q10	0.7009	
Q11	0.7148	
Q12	0.6623	
Q13	0.7325	
Q14	0.6467	
Q15	0.7138	
Q16	0.7810	
Q17	0.7029	
Q18	0.7806	
Q19	0.6582	

Figure 32. Analysis of Cronbach's Alpha for the Ease of Use Construct



Multivariate

Correlations

	Q20	Q21	Q22	Q23
Q20	1.0000	0.8333	0.6901	0.1491
Q21	0.8333	1.0000	0.8281	0.2609
Q22	0.6901	0.8281	1.0000	0.3858
Q23	0.1491	0.2609	0.3858	1.0000

Cronbach's α

	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.8219	

Excluded Col	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q20	0.7500	
Q21	0.6800	
Q22	0.6923	
Q23	0.9159	

Cronbach's α, standardized

	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.8153	

Excluded Col	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q20	0.7436	
Q21	0.6743	
Q22	0.6798	
Q23	0.9158	

Figure 33. Analysis of Cronbach's Alpha for the Ease of Learning Construct

Multivariate

Correlations

	Q24	Q25	Q26	Q27	Q28	Q29	Q30
Q24	1.0000	0.4667	-0.0559	0.1043	0.6180	0.0928	0.1936
Q25	0.4667	1.0000	0.2609	-0.4869	0.4944	-0.0928	0.2324
Q26	-0.0559	0.2609	1.0000	-0.1555	0.2764	0.3851	0.3464
Q27	0.1043	-0.4869	-0.1555	1.0000	0.0000	0.2752	0.5388
Q28	0.6180	0.4944	0.2764	0.0000	1.0000	0.4019	0.5745
Q29	0.0928	-0.0928	0.3851	0.2752	0.4019	1.0000	0.4749
Q30	0.1936	0.2324	0.3464	0.5388	0.5745	0.4749	1.0000

There are 1 missing values. The correlations are estimated by REML method.

Cronbach's α

	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.6792	

Excluded Col	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q24	0.6539	
Q25	0.6743	
Q26	0.6783	
Q27	0.7175	
Q28	0.5366	
Q29	0.6398	
Q30	0.5663	

Cronbach's α, standardized

	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.6832	

Excluded Col	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q24	0.6483	
Q25	0.6908	
Q26	0.6773	
Q27	0.7306	
Q28	0.5548	
Q29	0.6382	
Q30	0.5554	

Note: Question 27's α is larger than the set in both the normal and standardized.

Figure 34. Analysis of Cronbach's Alpha for the Satisfaction Construct

Multivariate

Correlations

	Q24	Q25	Q26	Q28	Q29	Q30
Q24	1.0000	0.4667	-0.0559	0.6180	0.1172	0.1936
Q25	0.4667	1.0000	0.2609	0.4944	-0.1172	0.2324
Q26	-0.0559	0.2609	1.0000	0.2764	0.3734	0.3464
Q28	0.6180	0.4944	0.2764	1.0000	0.3714	0.5745
Q29	0.1172	-0.1172	0.3734	0.3714	1.0000	0.4546
Q30	0.1936	0.2324	0.3464	0.5745	0.4546	1.0000

There are 1 missing values. The correlations are estimated by REML method.

Cronbach's α

	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.7151	

Excluded Col	α	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q24	0.7143	
Q25	0.6867	
Q26	0.7146	
Q28	0.5516	
Q29	0.7073	
Q30	0.6539	

Cronbach's α, standardized

	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Entire set	0.7267	

Excluded Col	Standardized	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Q24	0.7081	
Q25	0.7084	
Q26	0.7208	
Q28	0.5952	
Q29	0.7210	
Q30	0.6610	

Note: Question 27 was not included in this analysis.

Figure 35. Analysis of Cronbach's Alpha for the Satisfaction Construct with Redundant or Erroneous Items Removed

APPENDIX E. POST-TASK SCORING ANALYSIS RESULTS

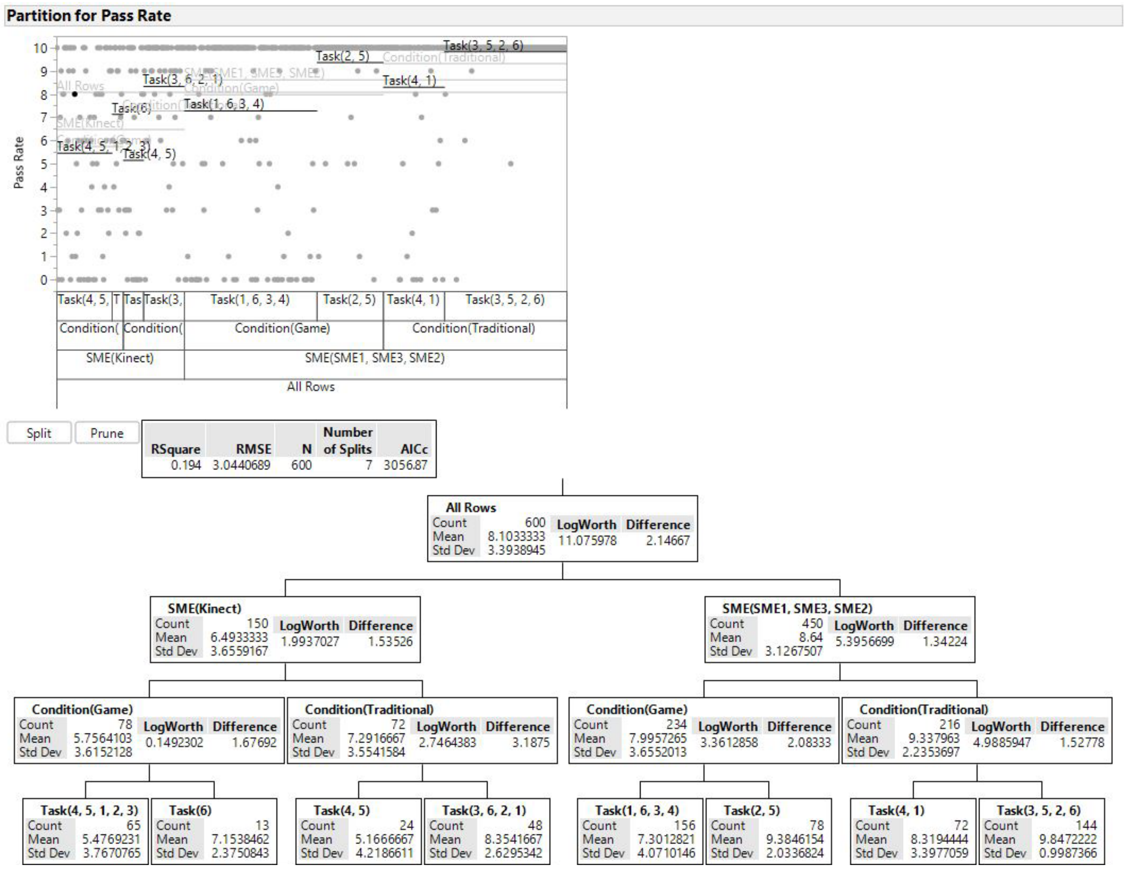


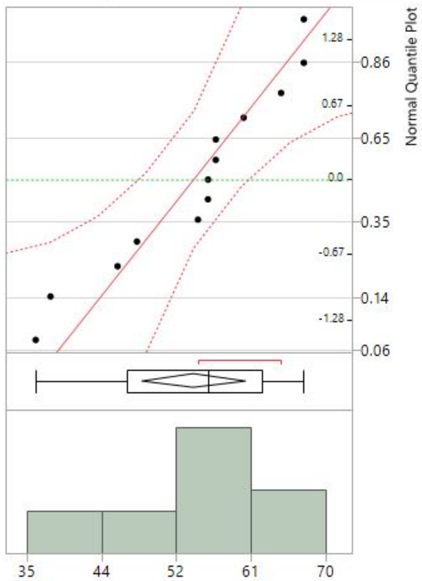
Figure 36. Partition Tree for Kinect and SMEs' Post-task Scoring

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX F. POST-TASK CONFIDENCE ANALYSIS RESULTS

Distributions Condition=Game

Overall Average



Quantiles

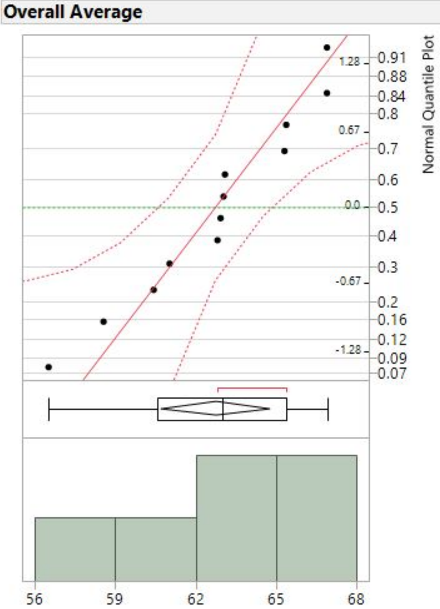
100.0%	maximum	67.06
99.5%		67.06
97.5%		67.06
90.0%		67.06
75.0%	quartile	62.2375
50.0%	median	55.911666667
25.0%	quartile	46.5
10.0%		36.525666667
2.5%		35.835
0.5%		35.835
0.0%	minimum	35.835

Summary Statistics

Mean	54.241795
Std Dev	10.100927
Std Err Mean	2.8014932
Upper 95% Mean	60.345724
Lower 95% Mean	48.137865
N	13

Distributions Condition=Traditional

Overall Average



Quantiles

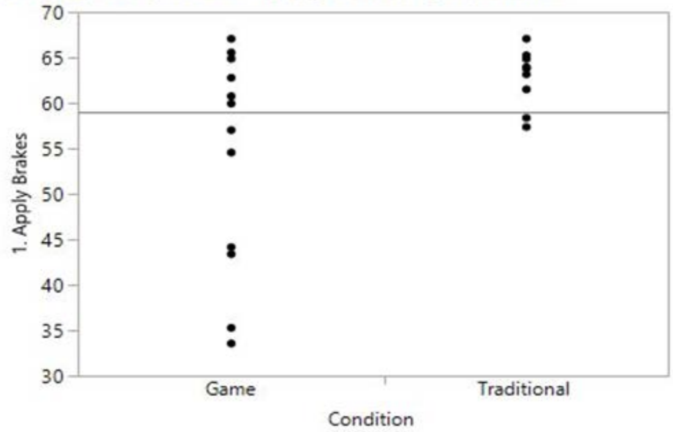
100.0%	maximum	67.06
99.5%		67.06
97.5%		67.06
90.0%		67.06
75.0%	quartile	65.47625
50.0%	median	63.023333333
25.0%	quartile	60.542083333
10.0%		56.9795
2.5%		56.345
0.5%		56.345
0.0%	minimum	56.345

Summary Statistics

Mean	62.772083
Std Dev	3.3056154
Std Err Mean	0.954249
Upper 95% Mean	64.872371
Lower 95% Mean	60.671795
N	12

Figure 37. Histograms, Box Plots, Quantile Plots, and Summary Statistics for Subjects' Overall Average Confidence Levels

Oneway Analysis of 1. Apply Brakes By Condition



Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
Game	13	139.000	169.000	10.6923	-1.608
Traditional	12	186.000	156.000	15.5000	1.608

2-Sample Test, Normal Approximation

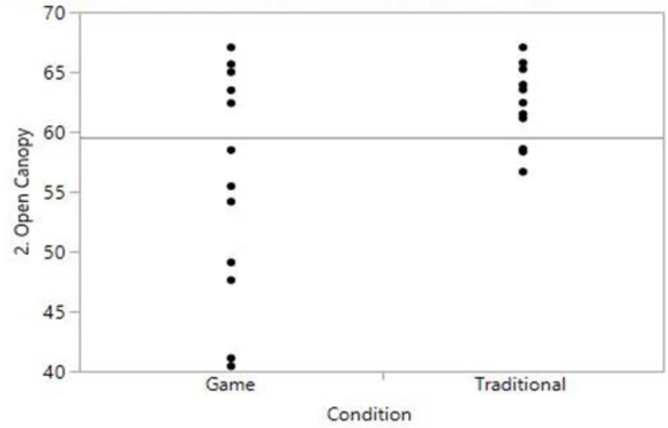
S	Z	Prob>|Z|
186	1.60768	0.1079

1-Way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
2.6730	1	0.1021

Figure 38. Wilcoxon Rank Sum Analysis for Subjects' Task 1 Confidence Levels

Oneway Analysis of 2. Open Canopy By Condition



Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
Game	13	143.000	169.000	11.0000	-1.390
Traditional	12	182.000	156.000	15.1667	1.390

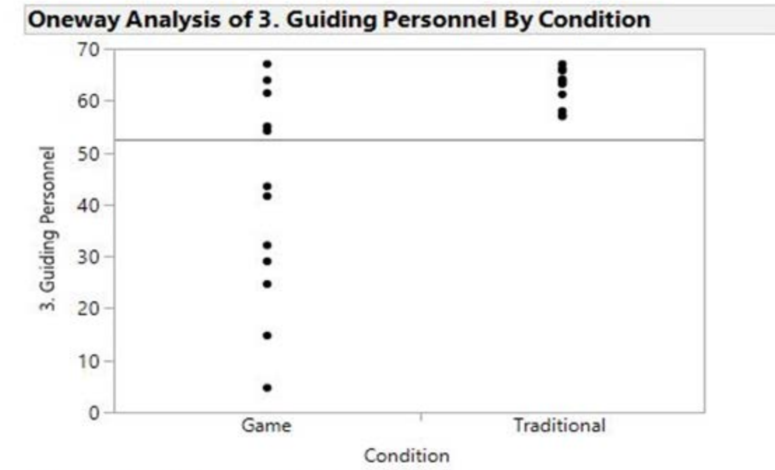
2-Sample Test, Normal Approximation

S	Z	Prob>|Z|
182	1.38969	0.1646

1-Way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
2.0077	1	0.1565

Figure 39. Wilcoxon Rank Sum Analysis for Subjects' Task 2 Confidence Levels



Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
Game	13	124.000	169.000	9.5385	-2.425
Traditional	12	201.000	156.000	16.7500	2.425

2-Sample Test, Normal Approximation

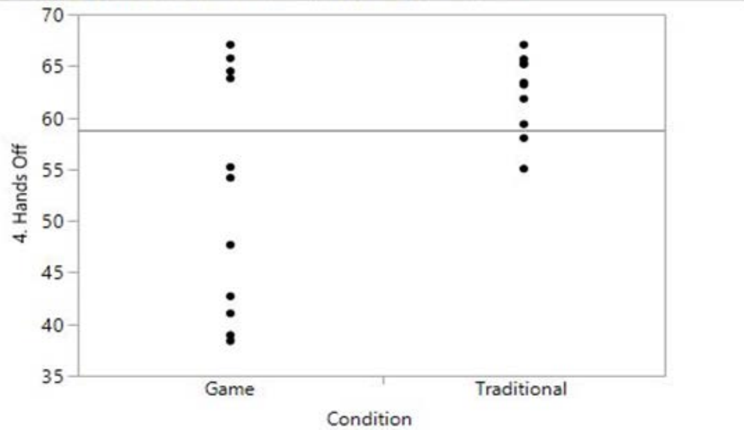
S	Z	Prob>|Z|
201	2.42515	0.0153*

1-Way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
6.0143	1	0.0142*

Figure 40. Wilcoxon Rank Sum Analysis for Subjects' Task 3 Confidence Levels

Oneway Analysis of 4. Hands Off By Condition



Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
Game	13	143.000	169.000	11.0000	-1.392
Traditional	12	182.000	156.000	15.1667	1.392

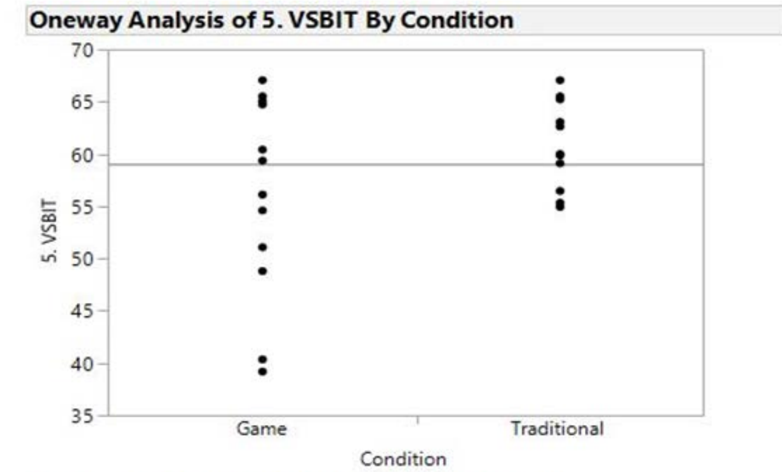
2-Sample Test, Normal Approximation

S	Z	Prob>|Z|
182	1.39238	0.1638

1-Way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
2.0155	1	0.1557

Figure 41. Wilcoxon Rank Sum Analysis for Subjects' Task 4 Confidence Levels



Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
Game	13	150.500	169.000	11.5769	-0.981
Traditional	12	174.500	156.000	14.5417	0.981

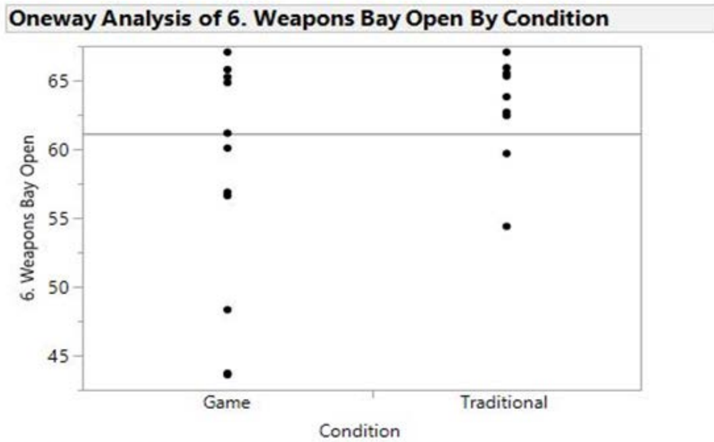
2-Sample Test, Normal Approximation

S	Z	Prob>|Z|
174.5	0.98115	0.3265

1-Way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
1.0169	1	0.3133

Figure 42. Wilcoxon Rank Sum Analysis for Subjects' Task 5 Confidence Levels



Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
Game	13	150.500	169.000	11.5769	-0.986
Traditional	12	174.500	156.000	14.5417	0.986

2-Sample Test, Normal Approximation

S	Z	Prob>|Z|
174.5	0.98573	0.3243

1-Way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
1.0264	1	0.3110

Figure 43. Wilcoxon Rank Sum Analysis for Subjects' Task 6 Confidence Levels

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Bernacki, M. L., Agular, A.C., & Byrnes, J. P., (2011). Self-regulated learning and technology-enhanced learning environments: an opportunity-propensity analysis. In G. Dettori & D. Persico (Eds.), *Fostering self-regulated learning through ICT* (pp. 1–26). Hershey, PA: Information Science Reference.

Bhardwaj, P. (2018, April 26). Two technologies could transform the way we all watch live sports—Here's how close they are to becoming reality. Retrieved from https://www.businessinsider.co.za/tribeca-film-festival-vr-ar-live-sports-2018-4

Bricken, M. (1991). Virtual reality learning environments: potentials and challenges. *ACM SIGGRAPH Computer Graphics*, *25*(3), 178–184. Retrieved from: https://dl.acm.org/citation.cfm?id=126657

Carson, E. (2015, April 8). 10 ways virtual reality is revolutionizing medicine and healthcare. Retrieved from https://www.techrepublic.com/article/10-ways-virtual-reality-is-revolutionizing-medicine-and-healthcare/

Ciavarelli, T. (2017, July 17). *Cognitive task analysis using GOMS*. Presented at the Naval Postgraduate School MOVES Institute, Monterey, CA.

Dempewolff, R. F. (1954). After disaster, the trouble shooter. *Popular Mechanics, 102*(3), 81–89.

Dettori, G. & Persico, D. (2011). *Fostering self-regulated learning through ICT*. Hershey, PA: Information Science Reference.

Eckstein, M. (2018, March 1). New simulators give naval aviators an edge in training for high-end fight. *USNI News*. Retrieved from https://news.usni.org/2018/03/01/new-simulators-give-naval-aviators-edge-training-high-end-fight

Headquarters Marine Corps. (2018, July 30). 2018 Marine Aviation Plan. Retrieved from https://www.aviation.marines.mil/

Helander, M., Landauer, T., & Prabhu, P. (1997). *Handbook of human-computer interaction (2nd ed.)*. Amsterdam: Elsevier.

Increasing adoption of AR and VR in gaming expected to drive growth of the global AR and VR market through 2025. (2017, May 22). *PR Newswire Europe Including UK Disclose*. Retrieved from http://search.proquest.com/docview/1900815062/

Kapp, K. (2012). Games, gamification, and the quest for learner engagement. *TD Magazine 66*(6), 64-68. Retrieved from https://www.td.org/magazines/td-magazine/games-gamification-and-the-quest-for-learner-engagement

Kirwan, B. & Ainsworth, L. (1992). *A Guide to task analysis*. Bristol, PA: Taylor & Francis.

Kolb, D. A. (2014). *Experiential learning: experience as the source of learning and development*. Upper Saddle River, NJ: FT press.

Lehman, A., O'Rourke, N., Hatcher, L., & Stepanski, E. (2013). *JMP for basic univariate and multivariate statistics: methods for researchers and social scientists*. Cary, NC: SAS Institute.

Mirt, J. (2017, December 14). 5 sports that are benefiting from virtual reality. Retrieved from https://www.viar360.com/blog/5-sports-benefiting-virtual-reality/

Moran, W. F. (2016, March 28). Assignment of U.S. Fleet Forces Command as executive agent for ready, relevant learning. Retrieved from http://www.public.navy.mil/bupers-npc/reference/messages/Documents/NAVADMINS/NAV2016/NAV16075.txt

Naval Air Systems Command. (2014, April 01). *Aircraft Signals NATOPS manual* (NAVAIR 00–80T-113). Coronado, CA: Commander Naval Air Forces.

Neller, R. B. (2017, February 7). Message to the Force 2017: "Seize the Initiative." Retrieved from https://www.albany.marines.mil/Portals/75/Docs/PAO/CMC-Message-to-the-Force-2017-Seize-the-Initiative.pdf

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Perlman, G. (2015, May 18). User interface usability evaluation with web-based questionnaires. Retrieved from http://garyperlman.com/quest/

Parnell, G. S., Bresnick, T., Tani, S. N., & Johnson, E. R. (2013). *Handbook of decision analysis* (Vol. 6). Hoboken, NJ: John Wiley & Sons.

Sall, J., Stephens, M. L., Lehman, A., & Loring, S. (2017). *JMP start statistics: a guide to statistics and data analysis using JMP* (6th ed.). Cary, NC: SAS Institute.

Stanković, S. (2016). *Virtual reality and virtual environments in 10 lectures*. doi:10.2200/S00671ED1V01Y201509IVM019

Tucker, S. (2009). F-35 Joint Strike Fighter Aircraft Systems Maintenance Trainer. Retrieved from https://www.textronsystems.com/sites/default/files/resource-files/aai_f-351109.pdf

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California