Spring 2009

The Effect of Whole and Part-Task Training and Feedback during Simulated Instrument Flight Training

Christian Jon Rossi

Embry-Riddle Aeronautical University - Daytona Beach

Scholarly Commons Citation

Rossi, Christian Jon, "The Effect of Whole and Part-Task Training and Feedback during Simulated Instrument Flight Training" (2009). *Theses - Daytona Beach*. 177.
https://commons.erau.edu/db-theses/177

THE EFFECT OF WHOLE AND PART-TASK TRAINING AND FEEDBACK DURING SIMULATED INSTRUMENT FLIGHT TRAINING

By

CHRISTIAN JON ROSSI

B.S., Embry-Riddle Aeronautical University, 2007

A Thesis Submitted to the
Department of Human Factors & Systems
In Fulfillment of the Requirements for the Completion of a
Master of Science in Human Factors & Systems

Spring 2009

THE EFFECT OF WHOLE AND PART-TASK TRAINING AND FEEDBACK DURING SIMULATED INSTRUMENT FLIGHT TRAINING

By

Christian Jon Rossi

This thesis was prepared under the direction of the candidate's thesis committee chair, Jon French, Ph.D., Department of Human Factors & Systems, and has been approved by the members of the thesis committee. It was submitted to the Department of Human Factors & Systems and has been accepted in fulfillment of the requirements for the degree of Master of Science in Human Factors & Systems.

THESIS COMMITTEE:

J. French

Jon French, Ph.D., Chair

Beth Blickensderfer

Beth Blickensderfer, Ph.D., Member

Barbara G Buck

Barbara Buck, Ph.D., Member

Shawn M Doherty

MS HFS Program Coordinator

Department Chair, Department of Human Factors & Systems

Richard H. Heist for JC

Associate Provost

ACKNOWLEDGEMENTS

The author wishes to thank all those who have helped on this project including those at the Boeing Company and Embry-Riddle Aeronautical University. Special thanks to Greg Atchison, Barbara Buck, and Cliff Sowadski for their help in creating the software and their overall guidance during the experimental process. Also, thanks to Dr. Jon French and Mike Gellar for all their hard work during the running of participants and the analysis stages of this experiment. This experiment required a lot of time and each of all of the aforementioned were very dedicated to seeing this project completed.

Abstract

The most cost effective method for training novice pilot's instrument flight procedures has not been well investigated. Part-task scenarios condition and a whole task scenarios condition, both with different levels of feedback were compared in a between group study of novice pilots learning a 737 instrument approach with Microsoft ESP flight simulator. The two different types of training methods were evaluated after a series of training exposures by comparing the ability to pass a series of training scenarios along with a final test. This standard was created by 737 pilots, FAA regulations, and 737 instructors. These training methods were also evaluated on which style of feedback produces the best performance score in three sub conditions. The three sub conditions of feedback consisted of elaborative (hierarchal structured) feedback, fixed feedback (knowledge of results feedback), and no feedback. While no difference in the ability to train the pilots in a series of training scenarios was found for the type of training (part-task versus whole-task), fixed feedback demonstrated the best overall performance when compared against the other two feedback conditions.

TABLE OF CONTENTS

List of Tables

List of Figures

Introduction

Statement of the Problem

Since the beginning of aviation, there has been a dream held by the masses of being able to strap into an aircraft, and fly. This dream cannot be achieved without extensive training as demonstrated since the beginning of the history of flight with the Wright Brothers and today's structured training to obtain a pilot's license. Flight training has evolved tremendously from solely being one of the most dangerous trials a person can go through to the safety culture of today. With the growth of aviation over the past one hundred years, there are increasing needs for additional pilots and more efficient means of training. Adaptive training with desktop simulation is one of these methods being created to circumvent the issues in the airline industry and the military to create more qualified pilots in a short period of time. By using today's technology to improve a pilot's skills and abilities, the future of training can only progress.

Before desktop simulation can be used as an official training method, research into the effectiveness of methods of training must first be investigated and rigorous FAA standards must be applied. Different methods include the presentation of the information and tasks that the student will be given must be evaluated. Two different training presentation styles that have an influential history in learning a skill are that of part task or whole task training. Also, another issue is the type of feedback that is given to the student to facilitate learning. Feedback has been well researched; however, there is still no consensus on which type of feedback facilitates learning better during tasks such as instrument flight (Hummel, Paas, & Koper, 2006; Hummel & Nadolski, 2002; Morrison, Ross, Gopalakrishnan, & Casey, 1995; Kulhavy, White, Topp, Chan, & Adams, 1985; Ende, 1983; Trowbridge & Cason, 1932).

Using desktop PC based flight training would produce a huge cost benefit ratio for flight schools if effective training methods could be demonstrated. Embry-Riddle Aeronautical University (ERAU) in collaboration with The Boeing Company, Phantom Works, used Microsoft ESP, an open code version of Microsoft Flight Simulator to test both part-task vs. whole-task designs of training scenarios along with fixed vs. elaborative feedback cues. In a single setting under similar experimental conditions, the results are expected to demonstrate that a.) Part-task scenarios yield faster results to train a student to a specific set of standards and that b.) Students given elaborative feedback will be able to identify the areas that need additional focus since they are given feedback with increasing information if an error is not corrected, allowing the students to reach standards faster than no feedback or fixed feedback.

The specific goals of this project are as follows:

1. Which type of training scenarios, part-task or whole-task scenarios allows the student to reach proficiency to the FAA standards (Boeing defined) the most rapidly?
2. Does the type of feedback have an effect on the ability to gain the skills to perform at FAA standards?

History of Flight Training

In 1910, there were less than ten aviators that had the full qualifications to be considered skilled. These aviators were mainly used for exhibition flights for the Wright Brothers to demonstrate their newest technological marvel, the airplane. As more planes were being developed by the forefathers of flight, the first flight school was created in Montgomery Alabama and it literally was a “crash” course in flying. The planes at the time were both unpredictable and unreliable leading to the training to become inefficient due to constantly occurring delays. Flight training could not occur if winds were too high or if Orville Wright, the only flight instructor wasn't available. Many of these flights lead to crash landings and damage to the few aircraft that were available (Ennels, 2002).

As the military leaders began to see that airplanes could be used for military operations, concurrently the advancement of technology and of training began to flourish. By 1947, the introduction of jet power in the F-86 Saber guided the recently formed Air Force to transition from propellers to these new types of aircraft (Lorell, 2003). After training approximately 200,000 pilots during World War II, the military understood how to successfully train pilots increasing the odds of survival for novice pilots (Werrell, 2005). Yet, jet powered aircraft posed a new challenge for transitioning old pilots that were used to propeller powered aircraft and the new pilots that would be first introduced to flying jets. Large numbers of accidents occurred during the late 1940s and early 1950s, mainly with the older pilots of World War II that were set in their old ways. This led to investigations into why the training was not producing the same efficiency as previous training in propeller aircraft. Investigators noticed that the main difference was the lack of standardization in the aircraft. Training equipment used different instruments such as altimeters and airspeed indicator. No standardization was used between the trainer and actual aircraft such as using different units of measurements (knots and mph), or not having a two seat trainer. Once these issues were resolved, flight training progressed and the military was able to produce numerous amounts of qualified pilots that would later fight in Korea and Vietnam.

As commercial airlines began growing, flight training became a large concern for civilian operators as well. Today's major carriers are mandated to give pilots proficiency training twice a year, if the pilot is given a new position which includes changes in a home facility, fleet types, or cockpit seat. Typically 15-20% of pilots will change position during a bid cycle, which consists of the pilots being able to put a request in for a new position (Xiangton et al., 2004). During this training, the pilots are paid but they are not providing any service to the airline, thus costing the airlines a great deal of money. If these pilots can be trained at a quicker pace, less of a cost due to downtime of pilots would be noticed. The classes that each pilot must go through have a specific length in time and each pilot must have a specific number of days off during this process (Xiangton et al., 2004). If the length of time of the classes can reduced, the instructors will be able to use their time more efficiently as well as having the ability to free up the flight training devices (FTDs) to be used by outside customers, resulting in a greater profit (Xiangton et al., 2004).

Simulation used in Flight Training

Gange (1954) defined a training device as a piece of apparatus used for training a student. Moreover, Gange (1954) explained that a simulator is a type of training device that has high fidelity to the actual system for which a student is training. Fidelity is a term that does not have an agreed upon definition, but can be best described as the ability for the simulator to closely replicate the actual environment for training purposes (Alessi, 1988). For a high fidelity simulator, it must resemble the actual system in operational equipment, display, controls, and the way one is affected by the operation of the system (Gange, 1954). Both training devices and simulators have been useful throughout the history of training students in the skills and abilities to become proficient at flight. When developing a simulator or training device, it must be understood that it should not just measure a student's performance, but it should also be designed to improve a student's performance.

Fidelity is an issue that comes into play with the design of simulators since the development of the first flight training device. Before describing the progression of how flight trainers/simulators were developed, one must understand what level of simulation is needed for a student to maximize transfer while decreasing the likelihood of confusion or stress (Alessi, 1988; Liu, Macchiarella, & Vincenzi, 2008). The common notion about simulation fidelity is that higher fidelity would result in a maximized amount of transfer and increased learning, but recent research has identified that this isn't always the case (Alessi, 1988; Liu et al., 2008). Liu et al. (2008) states that in current times, the level of fidelity needed for training is mainly concerned with cost-effectiveness. If a low amount of fidelity will produce the same results as a high fidelity simulation, why would one spend the money? Alessi (1988) described that the level of fidelity is dependent on the student's experience level. Students who are novices will have the best learning in low fidelity to medium fidelity simulators in comparison to a high fidelity simulator since a high fidelity simulator can cause confusion and become overly stressful resulting in little to no learning. Also, in less complex simulators, features can be removed allowing the student to attend to specific tasks which is essential for novices to learn procedural skills (Alessi, 1988). In comparison, students who are experienced may learn more in a simulator, but if they train in the actual aircraft, they may exhibit similar experiences of confusion as seen by the novice. High fidelity simulations and the actual aircraft (highest level of fidelity) are most effective for experts. Alessi (1988) comments that as students develop higher skills, the level of fidelity should increase since the complexity of the system increases, allowing for the student to progress based on their skill level as previously mentioned.

Liu et al. (2008) created a list of different aspects used to describe fidelity into seven parts; physical fidelity, visual-audio fidelity, equipment fidelity, motion fidelity, psychological-cognitive fidelity, task fidelity, and functional fidelity. Physical fidelity is identified as the simulator's physical properties such as visual-audio, equipment, and motion in relation to the actual system. Visual-audio fidelity is how closely the sights and sounds that the learner will see and hear during the simulation are to the actual flight. As technology increases, the ability for terrain mapping and better visual information also increases, which will reduce the amount of unexpected events for a student when using the actual system since the student will be fully immersed into a similar environment through the simulation. Equipment fidelity includes all software and hardware being used in the simulation. This was previously described as an issue with the military training devices having a lack of standardized equipment. If elements of hardware and software are not closely matched with the actual system, negative learning can occur, which resulted in many accidents during the military transition to jets from propellers (Werrell, 2005). Motion fidelity, the ability to reproduce similar sense of motion to the actual system, is an element that there is much debate on due to the effects being insignificant in most training (Liu et al., 2008). Psychological-cognitive fidelity is based on the ability of the simulation to reproduce similar psychological and cognitive demands placed on the learner allowing for similar stress and workload. Task fidelity is a result of the simulator requiring students to perform the same tasks in the simulation as if they were in the aircraft including inputs, reading instruments, and other responsibilities to successfully fly an aircraft. Lastly, functional fidelity is how the simulator reacts to the commands and tasks being performed by the student in regards to it acting the same as an actual aircraft. Each of these different aspects combined into a simulator will successfully allow for increased learning and transfer to occur (Liu, 2008).

The simulators and flight training devices used during the Wright Brother's time and even during World War I may look archaic in comparison to the simulators created after the digital revolution, but are not all considered to be low fidelity simulations. The first flight training device was achieved by students of the Wright Brothers, since weather conditions were unfavorable of flight. The students placed the airplane into the wind to obtain the effects of movement allowing for motion fidelity. One student would act as the pilot and another as a counterweight. By moving the elevators and stabilizers, the airplane would move, allowing for motor skill training to occur (Ennels, 2002).

Moving onto 1929, Edwin Link became the first person to file a patent on a flight trainer (Kelly, 1970). Link developed the Link Trainer due to the increasing cost of becoming licensed to become a pilot. The cost was a result of many instructors not wanting a student to solo since the aircraft was their personal property and the rarity of instructors in the 1920s. After utilizing a method called the "penguin system" developed by the French in World War I which involved practicing procedures on the ground in the aircraft, Link wanted to be able to create something that can be used anywhere and have the feel of flight in the controls. Previously, the "penguin system" was noted as creating top notch pilots with a lower amount of time and decreased cost since preliminary flight instruction was completed solely on the ground. Training on the ground allowed for less mechanical problems that consistently occurred with a decrease cost in operational cost since it was not in the sky burning a lot of fuel. The simulator operated by having an electric motor and air pressure to act similarly to a plane taxiing across an airfield while allowing for a simulation of roll, pitch, and yaw movements. The Link Trainer is now considered the first flight trainer that would allow for students to use simulation in a way that was never imagined before. At the time, the Link Trainer was considered to be one of the highest fidelity flight simulators available.

With the invention of the early computer, training devices obtained had dynamically moving instruments based on the user's input, along with strip charts for the instructors to study the performance of the student (Koonce & Bramble, 1998). Once computers turned from analog to digital, more advances were seen in the creation of training devices including new displays that allowed for color and scenery manipulation. Koonce and Bramble (1998) explain that while most early simulators were mainly for military use only, commercial airlines eventually utilized this technology to train their students as well. Although there was a large overhead for the commercial airlines to purchase simulators, the simulators eventually aided in decreasing training costs since students did not need to train on an actual aircraft (Dennis & Harris, 1998). Training on an actual system causes downtime of a system that should be in use to produce capital. Smode, Hall, and Meyer (1996) explain that early studies of flight simulations resulted in a decrease in training hours needed in the actual aircraft to reach a specific skill level based on a set of standards. As processing speed increased exponentially, so did simulated visual systems. This guided the way to the Federal Aviation Administration to eventually allow certified simulators for training in addition to actual flight training in the air (Koonce & Bramble, 1998).

During the pilot shortages in the commercial airline industry, airlines decreased the entry-level requirements and hired pilots with fewer flight hours and less experience. Previous training curricula were becoming obsolete, directing the way to new programs such as the advanced qualification program (AQP) created by the Special Federal Aviation Regulation (Longridge, Burki-Cohen, Go, & Kendrea, 2001). Programs such as the AQP allowed for a decrease in the amount of training by using simulation, while ensuring that the pilot has had the essential knowledge, skills, and abilities to perform at a set level of standards. It was projected that airline pilot training in the United States could utilize high fidelity flight simulators for major tasks, but also will include the use of low fidelity flight training devices for other tasks that require less realism (Longridge, Burki-Cohen, Go, & Kendrea, 2001).

In 1981, IBM created the first PC, later resulting in a flight simulator initially marketed for gaming entertainment. Soon after, joysticks, yokes, rudder pedals, and throttles were developed to help market the game, giving "the virtual flight experience" to a person sitting behind their desktop computer. With the increase in computing ability for home use, these games have now evolved into personal computer training devices which are more advanced than the training simulators used in the 1970s by the military and commercial airliners (Koonce & Bramble, 1998). Today, the FAA allows Part 141 also known as a FAA licensed training institution to use personal computer pilot training devices as long as they contain a specific set of standard features (Koonce & Bramble, 1998).

While there are inadequacies inherent in desktop simulators such as field of view restriction, confinement to front windscreen, aircraft instrumentation being nonstandard, the physics of the aircraft not being perfect, and the flight controls lacking a "real" feel, studies have shown that there are advantages in training procedural skills (Dennis & Harris, 1998). Ortiz (1995) demonstrated that desktop simulators helped students reach proficiency to standards in actual flight in a less amount of time (1.6 trials) than students who did not receive any training with a desktop simulator (3.1 trials). Also, desktop simulators allow for a student to have a preview of the pace of a particular task, assisting the development of a cognitive template of requirements necessary to achieve maximal performance without increasing workload (Dennis & Harris, 1998).

The application of desktop simulation is not necessarily confined to the classroom and can be utilized in offsite locations since many pilots have constraints on time and location availability. Distributed computer based training systems cannot replace the standard one on one training that one can receive through a human instructor, but it is considered to be the best medium of training when considering individual differences (Ramachandra, Cramer, Harville, & Ashworth, 2003). In a traditional classroom setting, teachers teach a multitude of students at a constant pace even though some students require different needs based on retention ability, years of experience, and hours needed to fully understand a concept. If experience is already at an acceptable level, inefficient usage of time and resources for a specific individual may occur (Ramachandra, Cramer, Harville, & Ashworth, 2003).

Longridge et al. (2001) established that instructors using high fidelity simulators must allocate their time between running the simulator, simulating radio communication, instructing and observing. When compared to actual flight training in an aircraft, instructors state that workload is much higher for training and running the simulator, leading to instructors needing more time to teach when using these devices (Longridge, Burki-Cohen, Go, & Kendrea, 2001). Since there are many constraints when dealing with the availability of high fidelity simulators and instructors, desktop simulators may be used to help circumvent the basic tasks needed to be learned, while lessening the amount of time an instructor is needed.

By incorporating an embedded instructor into desktop simulation, a student can perform tasks specific to their own needs with the guidance of a pedagogical agent (Magerko, Wray, Holt, & Stensrud, 2005). Magerko et al. (2005) describes an embedded instructor as an intelligent agent that is designed into the software to modify the training environment based on skill levels and inputs place by the student. An intelligent agent that adapts to the student will increase learning and comprehension of a task while maintaining interest by the student. Embedded instructors can guide a student on a path within the virtual environment necessary to obtain a certain set of skills and knowledge essential for a task, such as the procedural skills needed for flight. As the skills of the student increases, the embedded instructor should make fewer corrections, changes, or warnings to the environment and student (Magerko, Wray, Holt, & Stensrud, 2005). At the end of training, the embedded instructors also should allow for the removal of some variance due to instructor bias when evaluating training performance (Koonce & Bramble, 1998). The student's performance can thus be recorded and evaluated through a set standard without personal bias becoming a factor which can hinder training performance and the time needed for training.

Another element that will help develop the ability for desktop simulation to facilitate learning of knowledge, skills, and abilities to complete and perform tasks is use of scenario based training (SBT) (Oser, Cannon-Bowers, Salas, & Dwyer (1999). Oser et al. (1999) defines SBT as effective learning environments that allow for practice in a simulated environment that is similar to the actual conditions one will operate a system in while also receiving feedback based on events throughout the training. SBT contains four main components that show its training effectiveness as described by Oser et al. (1999); structured training, links between the tasks/learning objectives/performance measurement/feedback, empirical testing in many environments, and improved performance in team training. In SBT, a scenario uses different events and tasks (can focus on a single task or multiple tasks) to train the user and the results are collected and evaluated by a performance measurement tool based on the inputs of the user. The system must allow the student to make their own decision, but if errors are committed, feedback is given and the system, helping to control the path to intended learning objectives (Oser et al. 1999).

Currently, open source desktop flight simulators have been created by companies such as Microsoft, enabling researchers to develop scenarios that incorporate embedded instructors and complex tasks through open source software development kits (SDK). Microsoft ESP is one of these products recently released in early 2008, allowing companies to develop scenarios and eventually market them as commercial off the shelf training devices for commercial and military use. Along with the backing of research results indicating that desktop simulation can improve performance, products such as these can lead to the future of training by providing a low cost and easy to use environment with a level of fidelity that is constantly increasing based on the growth of technology. Due to open source coding ability, SBT can be effectively created into these serious gaming tools to allow for effective training to occur through desktop simulation.

Part-Task versus Whole-Task Approaches

In developing simulation scenarios for streamlining a student's training and performance levels, different methods have been researched in a large number of studies (Wightman & Sistrunk, 1987; Gopher, Weil, & Siegel, 1989; Kramer, Larish, & Strayer, 1995; Goettl & Shute, 1996). Two of these methods include part-task and whole-task training approaches (Wightman & Sistrunk, 1987; Gopher, Weil, & Siegel, 1989). The whole-task training approach consists of training a student in all aspects of a task throughout the entire process without any emphasis on specific aspects of the system they are training on (Gopher, Weil, & Siegel, 1989). The student is immersed into the system and asked to complete the task without any structure of increasing difficulty, responsibility, or scheduling. Whole-task training allows the student to develop strategies toward completing a task, but these strategies may be limited. Whole-task training also may be overwhelming for the learning depending on the difficulty of the task (Kramer, Larish, & Strayer, 1995). Since whole-task training is known for forcing individuals to adapt to complex and high workload situations with limited knowledge of the efficiency of their resources, many students do not explore further possible routes to solving a problem (Gopher, Weil, & Siegel, 1989). By decomposing whole-task elements of training into task parts, workload can be reduced and students can have the opportunity to explore these different pathways of attention strategies to reach the main goal and complete the task appropriately (Gopher, Weil, & Siegel, 1989; Kramer, Larish, & Strayer, 1995).

Part-task training methods can be broken into three subsets; segmentation, fractionation, and simplification (Gopher Weil, & Siegel, 1989; Wightman & Sistrunk, 1987; Goettl & Shute, 1996). Segmentation refers to breaking up a task into specific pieces on the basis of temporal or spatial dimensions. This allows for the student to practice tasks in isolation during full exposure to the main task before moving onto the whole-task. An example of this was performed by Wightman & Sistrunk (1987) by breaking a carrier landing task into starting positions of 6,000, 4,000, and 2,000 feet away from the runway, in which the error tolerances became more stringent when the distance was closer to the aircraft carrier. Wightman & Sistrunk (1987) assessed segmented part-task training through the segmentation of approach length during a simulated carrier landing final-approach task. By segmenting the approach into different lengths of distance until landing, Wightman & Sistrunk (1987) found that it improved performance during a final test in comparison to training with the whole-task training approach. Fractionation part-task training alternatively isolates tasks from the overall task and breaks down each task into specific elements to be practiced on their own (Gopher, Weil, & Siegel, 1989). Determining the elements to break down for training is difficult since isolating parts could possibly cause negative transfer and the skills practiced in isolation may not be integrated into the whole-task. Even though this can occur, when fractionation is performed correctly, it has shown more effective results in comparison to training with whole-task methods (Fabiani, Buckley, Gratton, Coles, & Dochin, 1989). One must understand which elements are important to the task and build the skills appropriately to lessen the chances of negative transfer. Lastly, simplification refers to reducing the difficulty of specific elements and gradually increasing the difficulty until it is equivalent to the whole-task (Gopher, Weil, & Siegel, 1989). One of the major criticisms of part-task training is that it encourages the development of learned behaviors that are unnecessary for completing the whole-task. When using part-task training, it is imperative the trainer eliminates any unnecessary, “deadwood” tasks associated with completing a part-task regimen (Goettl & Shute, 1996).

Whole-task training should highlight only tasks that have the highest priority during different phases of training (Kramer, Larish, & Strayer, 1995; Gopher, 1996). Kramer et al. (1995) studied the differences between whole-task training and a hybrid of whole/part-task training named variable-priority learning which focused the student on the whole-task at hand, but emphasized specific parts of the task during the training. Kramer et al. (1995) found that when the priority of tasks change, learning the task is easier than when all tasks are of equal importance. In addition, participants who utilized variable priority training performed significantly better in mastering an alphabet-arithmetic task and monitoring skills than those in the whole-task condition.

A different hybrid method known as a hierarchical approach of part-task training utilizes automation of specific components while exposing the student to the whole-task or components of the whole-task. Making use of the highly studied computer game Space Fortress II known for research in the area of complex skill acquisition, specifically simulating a complex and dynamic aviation environment, Shebilske et al. (1999) found that automating specific components that are near the top of the difficulty and rating of necessity, the hierarchy caused effective training to occur. Also using Space Fortress II, Gopher et al. (1994) indicated that when pilots were given a hierarchical part-task approach, pilots performed better than the whole-task group and emphasis/variable priority training group when comparing scores in the game, but emphasis/variable training led to the best transfer.

Gopher et al.'s (1994) main goal with utilizing Space Fortress II was to study the transfer of training from a complex game to a real life situation such as flight performance at the Israeli Air Force flight school. Two different approaches were used when creating groups; an emphasis-change approach (the Emphasis Only Training/Variable Priority group) and a hierarchical part-task approach (the Full Training group). In the emphasis-change approach, participants practiced the whole game at all times while being given instructions and support feedback to focus attention on various aspects of the game during the different trials. Workload during this approach was considered to be at a full load throughout the game allowing for the participants to learn coping strategies. In the hierarchical part-task training approach, participants were shown segments of the game that gradually became more complex and this group later participated in the whole game/task. The hierarchical participants were given the same feedback as the emphasis-change group, in addition to verbal tips on recommended behaviors (which pertained solely to the game). A control group was also created in which participants trained in the whole task without any prior experience or feedback.

Space Fortress II consisted of controlling a spaceship, flying it away from mines that were chasing the spaceship (also destroying these mines), managing resources (missiles), and destroying a space fortress (which was located in the center of the screen). The part-tasks consisted of seven possible part games; aiming, control of ship motion, trajectory control, change of weapon systems and hit mines, dynamic mine handling, fortress kill, and full game with friendly mines. Assessments of actual flight training were based on a group of 8 flights in which climbs, descents, turns, and complex combined maneuvers were studied along with a 30-deg turn, 45-deg turn, and departures from practice areas. The Full Training group (FT) that utilized the hierarchical part-task training approach scored significantly higher on the final game scores in all measures in comparison to the Emphasis Only Training group (EOT), but both performed equally during actual flight tests. In comparison to the control group, the FT and EOT groups both performed significantly better in flight performance. Possible explanations for the increased performance during flight tests are that attention strategies were being developed during the computer game which transferred into flight situations and that the participants were able to learn about the importance of exploring alternative responses and attention strategies that can be effectively used in alternative situations (Gopher, Weil, & Bareket, 1994).

In a second study using Space Fortress II, Fabiani et al. (1989) also tested the difference between two different approaches to part-task training during high complexity and low task organization scenarios. These two approaches again include hierarchical training and integrated training noted as Emphasis Only Training by Gopher et al. (1994).

Both the integrated and hierarchical groups appeared to learn at a slower curve than that of the control group, but this was due to the groups being told to focus on specific aspects/tasks in the game, resulting in lower overall scores. The hierarchical group outperformed the integrated group which outperformed the control group when comparing different variables of performance measures in the Space Fortress game, once again indicating the superiority of hierarchical part-task training over whole-task training. In specific aspects of the game, such as ship control, the integrated group performed just as well as the hierarchical group. Both part-task training groups did not become dependent on feedback since high performance was still maintained when feedback was removed. All participants were able to perform the secondary tasks in isolation to the game, but when combined with the game, the integrated training group was more resistant to disruption than the other two groups. Individual differences resulted in participants with either low or high abilities. Participants with low abilities in the hierarchical group outperformed low-ability participants in the integrated group. These low ability participants in the integrated group performed worse than the low ability control group participants indicating that integrated training is detrimental or of no value for low-ability participants. Participants with high-abilities were able to excel in either part-task training approaches.

By determining the correct method of training that should be used, the amount of time that a student needs to participate in the training can be drastically reduced. Through analysis of a comparison of whole-task and a hybrid method of part-task training, one can determine which raises performance to a set level of standards in the shortest period of time based on the abilities of the students being trained.

Feedback/Cueing

After determining the specific method of training for a training simulation regimen, one must determine the type of feedback that a student will receive based on their performance. Feedback, as seen with the type of training method, can also lead students to explore alternative responses when it is presented in the appropriate manner. Kulhavy et al. (1985) indicates that providing feedback after a response will assist in the retention of corrected information of targeted material that is to be learned. By correcting inaccurate information that was acquired during instruction, feedback can aid in increasing performance on a task without having much of an effect on the accurate information that was already obtained by the student (Kulhavy, White, Topp, Chan, & Adams, 1985). When feedback is presented correctly, correct responses can be confirmed and reinforced, increasing the comprehension of the task without increasing active processing (Morrison, Ross, Gopalakrishnan, & Casey, 1995).

Since the early twentieth century, feedback has evolved from being seen in three specific aspects as a motivator or incentive aspect of performance based on Skinner's theories, a connection of responses to stimuli based on Thorndike's theories, and information that a student uses to evaluate responses. Today, feedback is still well studied and current thought that indicates that feedback is an important piece of information for task mastery (Kulhavy & Wagner, 1993). Physiologists have even broken down how verbal feedback impacts the brain specifying that feedback caused activation in areas of the inferior parietal and anterior cingulate cortices, which have been suggested or important to processing performance feedback information (Kawashima et al., 2000). With this unit of measure, feedback encourages students to correct errors, confirm correct responses, and help students adapt in many situations (Kulhavy & Wagner, 1993). Some of the earliest research of feedback performed by Trowbridge and

Cason (1932) shows how feedback facilitates a student's ability to improve performance in these means. Trowbridge and Cason (1932) instructed students to draw a 3, 4, 5, or 6-inch line on a piece a paper while feedback was given to the subjects with various forms of auditory communication. The subjects were blindfolded throughout the entire experimental period and were never shown any of the experimental apparatus. The subject's left hand was stationary at all times during the experiment in order to improve the ability of the subject to draw a consistent line many times. Four different conditions of feedback were given: blank, nonsense, right-wrong, and correct feedback. Each subject drew 100 lines in one condition and was given a break before being placed into one of the other three conditions, where they would draw 100 additional lines. In the blank condition, no feedback was given to the students. Nonsense feedback indicated that the subjects would hear nonsense syllables whenever the subject finished a line. This feedback had no indication of if the line was the correct or incorrect length. Right-wrong feedback allows for the subjects to hear a statement of "Right" if the line was within 1/8 inch of a 3-inch line or "Wrong" in all other cases. Lastly, the correct condition allowed for the experimenter to indicate to the student a plus or minus score based on inches when the subject finished the line. This condition simply explained to the subject what they have done, performance wise. Trowbridge and Cason (1932) found without knowledge of correct feedback, subjects frequently did not know what they were doing a large portion of the time. Yet, when given correct feedback, subjects were able to formulate proper changes and adjustments when performing the same task on a new trial.

This early form of research on feedback helped open the door to categorize feedback into groups including cueing. Cueing is defined by Hummel et al. (2006) as an instructional technique that helps construct a representation or schema of problem-solving to be used in an assortment of diverse tasks. When cueing uses task-valid cognitive feedback, which includes information about task execution and characteristics, there is an enhancement in the ability to learn and improve in performance (Hummel & Nadolski, 2002). As with most forms of feedback, cueing has been shown to be most effective when presented immediately in comparison to delayed feedback (Hummel, Paas, & Koper, 2006; Hummel & Nadolski, 2002; Ende, 1983). Ende (1983) instructs that feedback and cueing that is given just-in-time should offer insight into the task, along with what consequences will occur through non-judgmental means. If feedback or cues are not given, mistakes may go uncorrected, correct actions are not reinforced and overall competency can decrease, lessening the ability for the problem solving schemas to emerge. As a result of presenting feedback as close as possible to the work that has been executed, students will be able to make the correct judgment about the specifics of what the positive and negatives of their actions were without coming to their own possible erroneous conclusions (Ende, 1983). Presenting immediate feedback information may cause issues such as being intrusive, but by using an acoustic modality, the possibility of perceptual and cognitive overload is decreased in comparison to visual feedback where information is displayed on a visual screen, which may overlay on top of necessary information needed for the task (Narciss & Huth, 2002).

Another aspect of how cueing should be presented is that it should serve as an embedded support device. By using an embedded tool such as an embedded instructor in simulation, cueing can give direction to a problem-solving process (Hummel & Nadolski, 2002). Cues and feedback in general must facilitate cognitive transfer by representing the task, while also allowing for applicability to a large range of tasks (Hummel & Nadolski, 2002; Kulhavy, et al. 1985). By having conditions in which students are able to understand the subject content, students will use the feedback more efficiently and can later apply this feedback to other related tasks (Kulhavy et al., 1985). The cues presented by an embedded instrument during training are best utilized when developed as corrective feedback (Waldersee & Luthans, 1994). Corrective feedback informs the learner when an error is occurring or has occurred which has been shown to enhance performance during training in comparison to the use of only positive feedback which does not support the ability to improve performance since it is only an indication that performance expectations are being met (Waldersee & Luthans, 1994). Waldersee and Luthans (1994) found that positive feedback groups tested in a customer service experiment resulted in significantly less accurate performance than those given corrective feedback. This finding is inconsistent with behavioral theoretical views since it indicates that positive feedback did not act as a reward system to improve and reinforce correct responses, but actually diminished it at times (Narciss & Huth, 2002). By giving the subject information about what improvements are needed to be made or the errors that were being committed, subjects were able to improve performance on the task (Waldersee & Luthans, 1994).

As with feedback, cues being used as immediate constructive feedback can be broken down once again into different categories, such as knowledge of results, knowledge of correct response, answer until correct and elaborative feedback (Narciss & Huth, 2006; Narciss & Huth, 2002; Morrison, Ross, Gapalakrishnan, & Casey, 1995; Corbalan, Liesbeth, & Merrienboer, 2008). Research on which category of feedback results in the best performance and learning has exposed many subtleties which lead to much confusion. Although well studied, the research has resulted in inconsistencies of opinion on which type of feedback is the best for improving performance and streamlining learning. The first category, knowledge of results simply indicates to the learner that their response was either correct or incorrect, usually by stating "Right" or

"Wrong" as performed in early studies such as Trowbridge and Cason (1932) (Narciss & Huth, 2006; Narciss & Huth 2002). Many studies have shown that this information can cause an increase in performance over conditions that give no feedback, but also have shown to confuse learners since they may not understand how to fix the error or problem (Trowbridge & Cason, 1932). This form of feedback can also be considered as a type of confirmation feedback since it confirms the correct response or indicates to the learner when an error has been committed, but does not indicate how to fix it (Schimmel, 1988). Narciss and Huth (2002), Kulhavy et al. (1985) and Schimmel (1988) point out that providing a small amount of information is most efficient for training new students. When adding any more information such as an explanation of why errors may have occurred, feedback can be considered distracting or no more useful than presenting the minimal information. This is indicative of the many differences in opinion on feedback between researchers when comparing knowledge of correct response and elaborative feedback. When selecting additional information to present as feedback or a cue, it must be done carefully to help foster accurate processing of the information that is given to the learner (Narciss & Huth 2006.)

The second category of feedback, knowledge of correct response, provides the correct answer to a situation, independently of the correctness of solution steps performed by the learner (Corbalan, Liesbeth, & Merrienboer, 2008; Narciss & Huth, 2006; Narciss & Huth 2002). Corbalan et al. (2008) showed that this category can provide the steps to complete a problem or by demonstrating the solution step by step; this type of feedback increases learning in comparison to only providing knowledge of results. Also, in comparison to the third feedback category, "answer until correct", Morrison et al. (1995) demonstrated that knowledge of correct response yielded better performance in a computer based instruction task that involved learning different information through computer lessons and answering questions in a post test about what they have previously learned. "Answer until correct feedback" requires that students act in response to an error until the task is completed (Narciss & Huth, 2006; Morrison, Ross, Gopalakrishnan, & Cason, 1995). "Answer until correct feedback" has also produced negative results in numerous studies than that of control groups in which no feedback was presented (Morrison, Ross, Gopalakrishnan & Cason, 1995). This is mainly due to the fact that many subjects presented with the "answer until correct type of feedback" result in guessing over numerous tries until they complete the task. This leads to a low efficiency in learning and is not recommended for training (Morrison, Ross, Gopalakrishnan, & Cason, 1995).

Similarly as demonstrated earlier with the methods of training, the most efficient means of providing information is by using a hybrid of feedback or cues as performed with the last category of feedback known as Elaborative Feedback (Corbalan, Liesbeth, & Merrienboer, 2008; Narciss & Huth, 2006; Narciss & Huth 2002). Elaborative Feedback identifies incorrect and correct responses while also providing more information detailing why incorrect responses were incorrect, without immediately identifying the solution (Morrison, Ross, Gopalakrishnan, & Casey, 1995). The additional information can include explanations for the answer, the type of errors that were committed, hints about procedural skills or sources of information, and strategies on how to solve the problem (Narciss & Huth, 2002). Narciss & Huth (2002) state that when trying to initially learn a skill, the feedback or cues should not immediately provide the solution or explain the correct strategy, but should provide stepwise manageable pieces that allow the student to try to fix the error multiple times and correct their errors with the least amount of feedback necessary. If all of the information is presented at once, there is possibility of cognitive overload and confusion as exhibited with knowledge of correct responses. Elaborative feedback is a type of informative tutoring feedback. Informative tutoring feedback's overall goal is to provide useful information without immediately presenting the correct solution and guiding students to detect their own errors and eventually correct them with limited assistance (Narciss & Huth, 2006). By breaking down feedback into three steps; identifying errors, providing guidance towards a correction, and offering praise for work, Schimmel (1988) found that the learning of intellectual skills increased along with the ability to replicate the information learned. When studying types of feedback, Schimmel (1988) found that informative tutoring feedback or more precisely Elaborative Feedback presented in this manner was the only type of feedback to show of the intellectual. However, identifying the stepwise process of presenting information is difficult and costly. Some researchers support the need for the cost, while others do not see a significant improvement, deeming the cost unnecessary, again exhibiting the inconsistency of views on feedback (Morrison, Ross, Gopalakrishnan, & Casey, 1995).

Overall, feedback should relate to the learning outcome and signify to the students what errors are being committed, eventually allowing the student to work through an issue and ultimately solve it. The information should relate to the topic, task, errors, or solutions (Narciss & Huth, 2002). Comparisons of different types of feedback are inconsistent, thus resulting in the need to investigate which category will facilitate the most efficient learning. Also, feedback cues are rarely examined in a flight training situations. This deficiency highlights the need to study if highly complex tasks like with high workload such as landing aircraft and instrument flights are well-matched with an informative feedback style such as elaborative feedback, or if a simple feedback technique such as "knowledge of results" is all that is needed.

Hypothesis

In the reasons elaborated in the introduction above, it is hypothesized that participants in the hierarchical part-task scenarios with elaborative feedback will be able to reach and maintain standards in comparison to all other conditions. Workload in this condition (part-task scenarios with elaborative feedback) is expected to be the lowest in comparison to the other three conditions. Also, detailed information on what errors are being committed through elaborative feedback is expected to cause participants to quickly identify problem areas requiring more focus. Specifically these can be stated as:

1. Part-task scenarios will result in more participants passing the final across all conditions.
2. Elaborative feedback will result in more participants passing the final across all conditions.
3. Part-task scenarios with elaborative feedback will result in more participants passing the final in comparison to all other combined groups.
4. No feedback will result in the lowest amount of participants passing the final (as this was a control group).
5. Workload will follow the above hypotheses:
    a. Lowest amount of workload seen in the:
        i. Part-task scenarios
        ii. Elaborative feedback
        iii. The combination of part-task scenarios with elaborative feedback
    b. Highest amount of feedback seen in the no feedback condition.
6. Training errors will follow the above hypothesis:

a. lowest amount of training errors for:
    i. Part-task scenarios
    ii. Elaborative feedback
    iii. The combination of part-task scenarios with elaborative feedback
b. Highest amount of training errors for no feedback.

Method

Participants

In this study, a total of 103 student pilots from Embry-Riddle Aeronautical University participated. Participants were required to hold a private pilot's license with less than 150 hours experience and could not hold an instrument rating. The participants included both males and female pilots. Participants were monetarily compensated at the completion of the study with $100 dollars. The compensation is not based on hourly commitments due to the fact that the time to complete the study varied based on performance. Participants were advised that they can withdraw from the experiment at any time with no risk, along with all other information necessary in a consent form. This study was approved by the university IRB committee. Participants were recruited through ads placed around Embry-Riddle flight departments, class meetings, and flyers placed in mailboxes of all students in the University. Participant performance was monitored by a performance assessment tool provided by The Boeing Company. An experimenter collected this information and was present throughout the experiment to answer questions.

Apparatus

Microsoft ESP, the commercial open source code SDK version of Microsoft Flight Simulator, was placed onto four desktop computers powered by dual-core processors and NVIDIA GeForce 8800 GT 512MB video cards (or equivalent). The desktop computers each had 2 monitors placed vertically above each other allowing for an inside the cockpit (static) display on the lower monitor and an outside window view (dynamic) display on the upper monitor. Each computer station was fit with a Saitek Pro-Flight Yoke, Saitek Throttle quadrants and a Saitek Rudder Pedal set. Participants were given noise canceling headphones which reduced the amount of distraction from outside noise. Each computer also contained the Boeing performance assessment tool that ran in the background during the simulation to capture second by second performance data during all the flights except familiarization periods. Throughout the scenario, Microsoft PowerPoint was utilized to display a hints file to identify tips on landing the aircraft. The PowerPoint file was displayed on top of the view of the cockpit instruments in-between flight scenarios. Experimental apparatus was placed at the Embry-Riddle Team Gaming Simulation Laboratory (Fig. 1, Fig. 2).

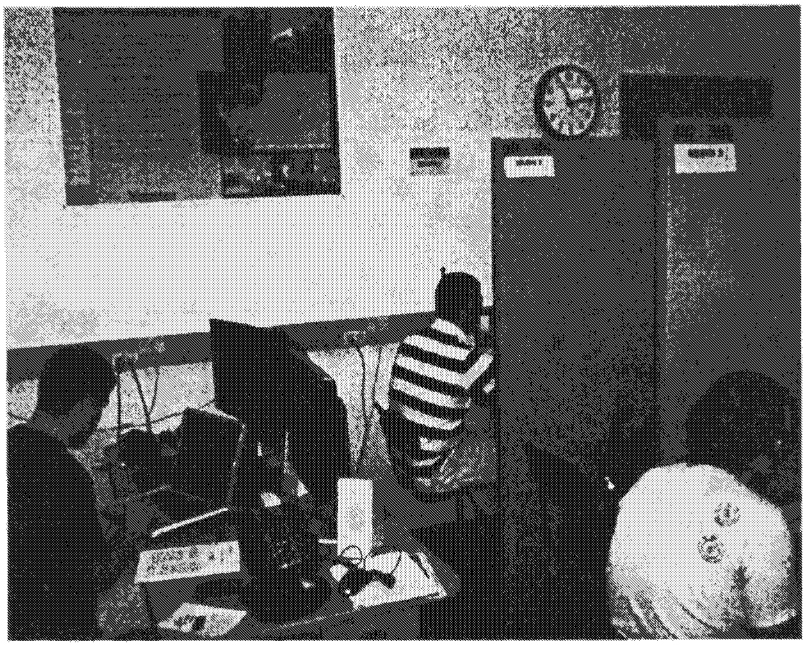

Fig. 1: Team Gaming Simulation Lab

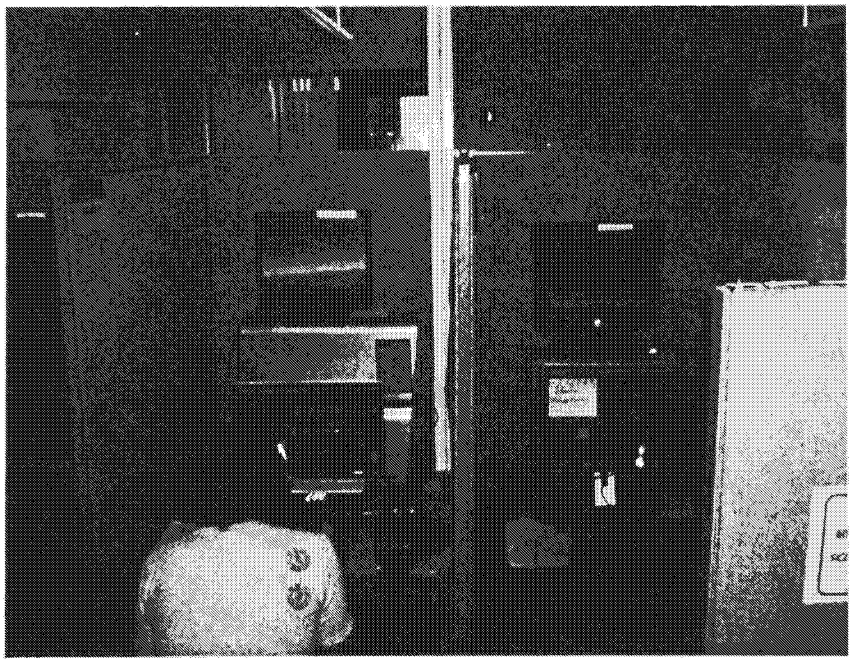

Fig. 2: Team Gaming Simulation Lab with Study Apparatus

Design

The design of this study consists of a 2x3 between subjects design. Participants were randomly assigned to each of the groups; part-task elaborative feedback, part-task fixed feedback, part-task no feedback, whole-task elaborative feedback, whole-task fixed feedback, whole-task no feedback (Fig. 3). All groups were exposed to a familiarization flight period consisting of three tasks; familiarization of the displays, controls, and flight dynamics of the 737-800 aircraft, embedded co-pilot walkthrough of the landing into Seattle Tacoma International Airport (KSEA) Runway 34R, and a demo landing with full control of the aircraft systems. The period was presented to train participants to reach a standard level of experience with Microsoft ESP and the Boeing 737-800 aircraft simulation. The period of familiarization lasted approximately 30 minutes per task (total time = 1.5 hours). No data from the performance assessment tool was collected during the familiarization period.

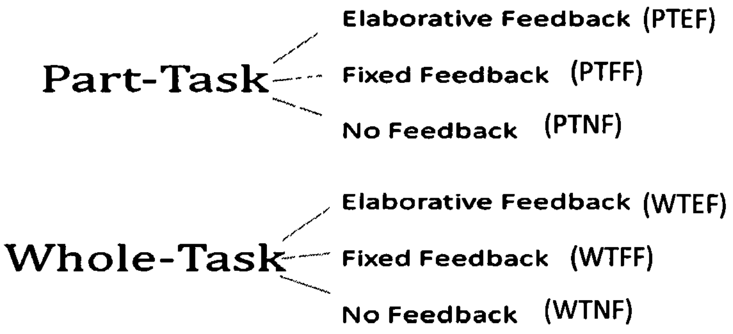

Fig. 3: 2x3 Study Design

At the completion of the familiarization period, participants were placed into the training segment which varied based on the independent variables; whole-task or hierarchical part-task scenarios and elaborative, fixed, or no feedback (Fig. 3). All groups were required to land a Boeing 737-800 aircraft into KSEA Runway 34R. They were given instructions from an embedded co-pilot on standard flight information such as when to make the turn, when they are approaching glide slope, etc. The flights consisted of starting straight and level at an airspeed of 210 knots indicated airspeed (KIAS), altitude of 4,000 feet, and a heading of 250 degrees. The student had to maintain airspeed by +/- 10 KIAS, altitude +/-500 feet, and heading +/- 10 degrees. At approximately 21 DME, the student was given instructions to make a turn to heading 341 degrees to line-up with the approach and immediately following their localizer at the completion of the turn. All performance measures must be maintained with an addition of bank angle target at 25 degrees, not exceeding 30 degrees. Once the turn was made, at approximately 17 DME, the embedded co-pilot made an announcement to descend to 2,200 feet. All performance measures must be maintained with the addition of descent rate at 1,500 feet per minute with and approximate +/- 500 feet per minute tolerance. During the descent, flaps must be placed down to 5 and airspeed reduced to 175 KIAS. At 11 DME, the embedded co-pilot made an announcement to configure the aircraft for landing. At this time, the aircraft must have the landing gear placed in the down position, flaps set to 30 degrees and airspeed reduced to 145 KIAS. At this point, the participant had to guide the aircraft to the runway still utilizing the localizer needle with the addition of glide slope indicators. At touchdown, the aircraft must land on the runway, reduce throttle to 0 KIAS, and engage brakes.

Whole-task scenario groups (all feedback levels) were exposed to the landing simulation with full control of the following flight systems; airspeed, altitude, steering, flaps, landing gear, and brakes. Participants had the ability to train to reach standards for up to 20 trials. If failure occurred during a trial, the participant retried the landing procedure from the beginning of the flight. Failures consisted of obtaining more than 10 warning cues, crashing, landing off the runway, and pre-specified boundaries. After one successful landing, the participant moved to a final test (total training time = approximately 1-3 hours).

Participants exposed to the hierarchical part-task scenarios method were given five levels of training based on levels of control. Automation of flight systems determined the level of control. The automation allows for the computer to take over control of the necessary flight systems during the landing procedure. During the lowest level of control (Level 1), the automation of the flight systems was the highest. Levels of control include:

- Level 1 – Steering and Limited Altitude (altitude only controlled after intercepting glide slope at 6.2 DME)
- Level 2 – Steering and Altitude
- Level 3 – Steering, Altitude, Flaps, and Landing Gear
- Level 4 – Steering, Altitude, Flaps, Landing Gear, and Brakes
- Level 5 – Steering, Altitude, Flaps, Landing Gear, Brakes, and Airspeed

Participants were initially placed in Level 1. The participant was given up to four tries to pass Level 1. If failure consistently occurred in Level 1, after the fourth trial, they were moved onto Level 2. If failure was once again consistent for four more trials, the participant was removed from the experiment and the training was considered a failure. This logic remained throughout each level of control, meaning that if a participant started a new level of control and fails 8 consecutive trials, they were removed from the study. If the participant was able to pass during any of the trials, they were moved to the next level of control. After reaching Level 5, when the participant was completed (either pass or fail 4 times, but passed level 4), they were moved to the final test (total task time = approximately 1-3 hours). Both hierarchical part-task and whole-task scenario groups' performance was recorded using the Boeing performance assessment tool provided by The Boeing Company.

In addition to being exposed to different types of scenario methods, the participants were placed into groups based on feedback presentation such as no feedback, fixed feedback, and elaborative feedback. This feedback was presented by the simulated embedded co-pilot who acted as a virtual instructor and gave verbal warning cues when an error or violation was committed. Elaborative feedback cues acted as a form of hierarchal cueing in that as more cues were presented, more information was given to the student to correct the problem. For example, if the student was consistently having errors with heading, the embedded co-pilot would give verbal cues of "Check Heading," followed by a second cue of "Check Heading 341," and a third cue of "Turn Right/Left Heading 341." If the student was able to correct the error after the first cue, the cueing terminated. In comparison, fixed feedback cues only state one consistent cue even if the student was committing the error for a long period of time. For example, the co-pilot would state, "Check Heading" multiple times until the heading error was corrected. Participants placed in the no feedback condition were not given any cues or warnings throughout the entire training period. The no feedback condition was placed as a control in this study to identify if exposure had an effect on the ability to pass the final test. Please refer to Table 1 for all feedback cues presented in this experiment.

Table 1: Feedback Cues

Action	Fixed Cues	Elaborative Cues
Heading	"Check Heading"	"Check Heading"
		"Heading 341"
		"Turn R/L heading 341"
Altitude	"Check Altitude"	"Check Altitude"
		"Maintain 2,000"
		"Climb/Descend and Maintain 2,000"
Air Speed	"Check Air Speed"	"Check Air Speed"
		"Maintain 210 knots"
		"Decrease/Increase Air Speed to 210 knots"
Flaps	"Check Flaps"	"Check Flaps"
		"Check Flaps Down"
		"Set Flaps at 5 / 30"
Landing Gear	"Gear Down"	"Gear Down"
		"Check Gear Down"
		"Pull Landing Gear Lever Down"
Course	"Course 341 degrees"	"Course 341 degrees"
		"Left/Right of Course"
		"Right/Left of course. Correct to the right/left"
Glide Slope	"Check Glide Slope"	"Check Glide Slope"
		"Above/Below Glide Slope"
		"Above/Below Glide Slope, Climb/Descend"
Engage Brakes	"Brakes"	"Brakes"
		"Engage Brakes Immediately"
		"Engage Brakes Immediately by pressing pedals"
Turning Bank Angle	"Check Bank Angle"	"Check Bank Angle"
		"Bank Angle too high"
		"Decrease Bank Angle turn right/left"

After the completion of training, all participants were required to fly a final test mission. The final test was the same mission as performed in training, but without any feedback and full control of the aircraft (total final test time = .5 hours). At the completion of the final test, participants were required to immediately fill out a subjective workload assessment questionnaire (NASA-TLX paper and pencil) considering the workload throughout the entire training and final test.

Dependent Measures

Performance assessment. Utilizing the proprietary performance assessment tool created and provided by The Boeing Company, performance measures including second by second data of altitude and speed were collected along with information of performance tolerance deviations and violations for heading, altitude, airspeed, course, glide slope, bank angle, flaps, landing gear, and brakes. The information of tolerances was based on the feedback cue presentation throughout each trial of training and the final test. The data utilized was pass/fail on the final which is essentially a whole-task no feedback flight (final instrument approach) in a 737 into simulated SEATAC (KSEA) airport. The criterion for passing the final was to successfully land the 737 within FAA standards for instrument landing and while accumulating less than 100 violation points as defined by the Boeing Company parameters for a successful flight. Other measures of flight performance were collected, but the pass/fail measure was primary. Due to the pass/fail nature of the dependent measures, unequal N's per group, non interval data (ordinal), violation of homogeneity of variance, and small sample sizes nonparametric comparisons were used for analysis purposes. Tests such as the Kruskal-Wallis and Mann-Whitney *U* tests were used to evaluate for all results with an alpha level of $p<.05$ (one-tailed).

Workload assessment. When administering training methods, workload, the expense of a human to attain a criterion performance, was a necessary measure for understanding if participants were both physically and cognitively overloaded throughout the learning process. If workload was excessive, it can lead to physiological or psychological stress or adaptation to a lower criterion of performance (Hart & Staveland, 1988). The NASA-Task Load Index (NASA-TLX) is a highly administered workload assessment tool that has been utilized for the past 20 years based on measurements of mental, physical, and temporal demands, frustration, effort, and performance (Hart, 2008). Since its creation in 1988, thousands of studies have utilized this assessment tool to measure workload, while many others use it as a benchmark to test the efficacy of theories and models (Hart, 2008). Through the study of subjective experience of workload, one can find the differences between the specific demands placed on a person from the task and the responses of each participant to those demands (Hart & Staveland, 1988). If a specific part of the task was affecting workload negatively, counter-measures can be made in the future to alleviate this strain. Due to the NASA-TLX being a likert scale, a nonparametric Kruskal-Wallis comparison was used for analyses.

Procedure

When initially arriving to the lab, participants met in groups of up to 4 and were given instructional packets that contain all information needed for the experiment. The packet included experience questionnaires, the consent form, directions based on their group, NASA-TLX paper and pencil workload assessment forms, out briefing information, and contact information. All documents in the packet can be found in Appendix A. The instructor first read the consent form to the participants and asked for a signature.

Once the previous steps were completed, participants filled out a questionnaire based on their flight and gaming experience. Other information in this form pertained to their demographics. Refer to Table 2 for all information on participant demographics. More detailed information on demographics can be found in Appendix B. The number of total flight hours for each participant was collected since they were required to have no more than 150 total flight hours logged, along with holding a private pilot's license. Figure 4 shows that there are similar total hours logged throughout each group.

Table 2: All demographic information by Group. Flight hours indicate the actual Flight Time logged by each participant in an aircraft. Based on flight training at Embry-Riddle Aeronautical University, most of this time for participants is in a Cessna 172 SP.

	Demographic	PTEF	PTFF	PTNF	WTEF	WTFF	WTNF
	Number of Participants	19	17	15	18	18	16
	Average Age	20 (1.8)	22 (5.1)	20 (1.8)	20 (1.6)	20 (2.1)	20 (1.7)
Gender	Number of Males	17	16	14	17	18	14
	Number of Females	2	1	1	1	0	2
Flight Hours	Total - Average	91.9 (29.7)	98.1 (29.2)	95.7 (22.3)	92.0 (23.5)	88.8 (30.5)	103.8 (23.1)
	Last 12 Months - Average	48.7 (36.4)	45.7 (38.3)	46.2 (23.7)	54.9 (29.9)	45.9 (29.1)	52.8 (27.5)
	Last 90 Days - Average	7.9 (10.2)	20.4 (25.3)	12.0 (12.3)	17.8 (19.7)	14.8 (11.2)	10.5 (10.7)
Video Game Flight Sim Experience	0 hours	3	2	0	0	1	0
	0-100 hours	10	13	9	14	13	10
	100-500 hours	3	2	3	3	3	3
	500-1000 hours	1	0	2	0	1	2
	1000+ hours	1	0	1	1	0	1
737 Sim Experience	No Experience	11	12	8	13	11	7
	<15 hours Experience	7	4	2	1	3	7
	≥15 hours Experience	0	1	4	4	3	2

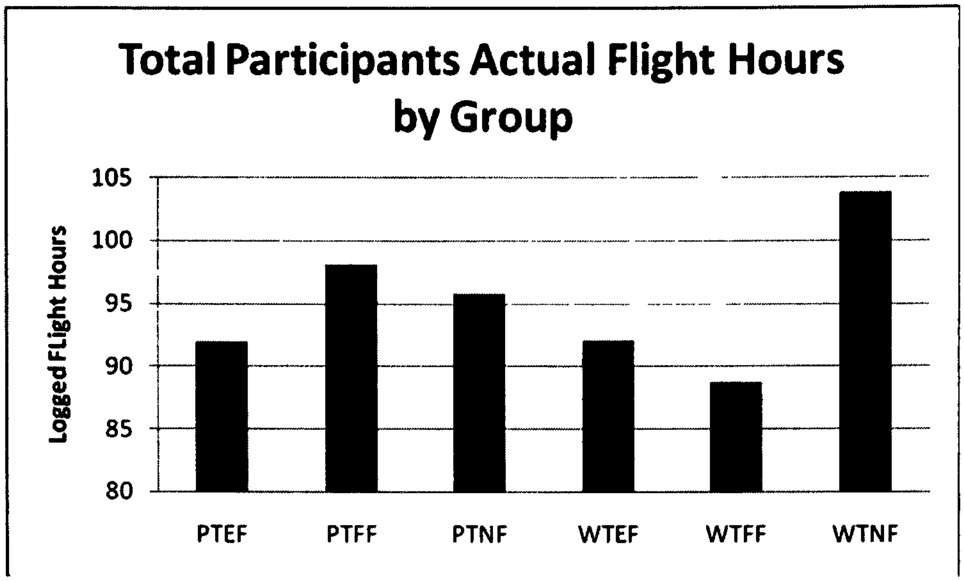

Fig. 4: Total Logged Flight Hours by Group. This identifies that each group had similar logged flight hours in an actual aircraft. Error bars show the standard error of the mean for each group.

The flight gaming information was collected in order to identify if the participants specific background experience influenced the results during the training and final test. Figure 5 shows the amount of previous flight simulation video game experience that the participants had based on their condition. As shown in Figure 5, most participants had 0-100 hours of video game flight simulation time before being part of the study. All groups had a similar range of video game flight simulation experience throughout the participant selection. Table 3 shows the amount of 737 video game flight experience (on their own time/before study) that participants had by group. This shows that each group once again had a similar range of 737 flight simulation experience, with most participants having no experience at all.

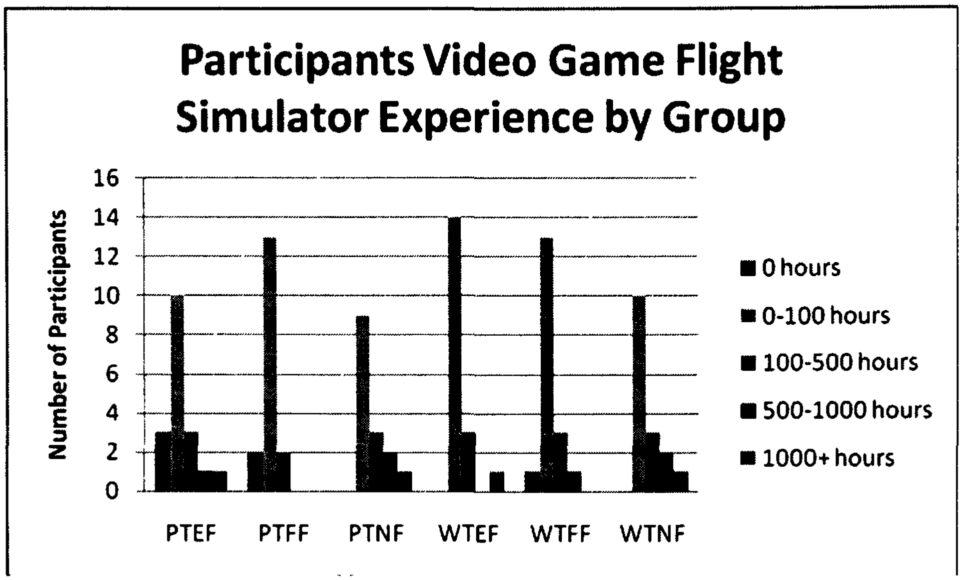

Fig. 5: Flight Simulation Experience by Group. This experience was based on the participant's previous exposure on their own time and was not an independent variable in this study. If hard to read, the legend follow left to right on the graph.

Table 3: 737 Experience throughout groups. This experience was based on the participant's previous exposure on their own time and was not an independent variable in this study. 737 experiences included Flight Simulation video games along with any other flight time in an actual 737 or Flight Training Device.

Group	None	<15 hours	>15 hours
PTEF	11	7	0
PTFF	12	4	1
PTNF	8	2	4
WTEF	13	1	4
WTFF	11	3	3
WTNF	7	7	2

Dependent on which group they were in, participants were read a set of instructions by the instructor that explains the familiarization flights, the training flights, and the final mission/test. When the forms were completed, the participants were notified that the task is very difficult and to keep trying since all data was anonymous. This notification gave motivation to the participants since it was needed due to the complexity of the task.

The participants were walked over to the computer simulation stations and shown the graphics placed on the side wall that displayed the cockpit layout. The participants were also instructed on how to use the trim (based on the gaming controls used) by the instructor. As previously explained, the students were now placed into the familiarization flight period. From this point on, the computer simulation was fully automated and the instructor watched to make sure that everything is running smoothly. After completing the three familiarization flights, participants were immediately placed into the training portion of the flight which included the approach and landing of the 737-800 aircraft. When the final task was completed the participant was removed from the computer and completed a NASA-TLX form. The instructor walked the participant through the rest of the packet that also included contact information for payment purposes. The participant was then debriefed about the overall study and thanked for their contribution to the study. In summary, the procedure was as follows:

1. Read and sign consent form
2. Complete background/experience questionnaire
3. Instructor reads all the instructions about the study
4. Instructor informs participants on difficultly of task and walks participant to the simulation computers
5. Familiarization flights
    1. How-To Fly
    2. Co-Pilot Walkthrough Landing
    3. Baseline Mission
6. Training session
    1. Seattle Approach
        1. Part-Task
            1. 5 Modules – Control of:

         1. Steering only (Heading)
         2. Steering and Altitude
         3. Steering, Altitude, Landing Gear, and Flaps
         4. Steering, Altitude, Landing Gear, Flaps, and Brakes
         5. Steering, Altitude, Landing Gear, Flaps, Brakes, and Airspeed
   2. Whole-Task
      1. 1 Module – Control of
         1. Steering Altitude, Landing Gear, Flaps, Brakes, and Airspeed

7. If training session is passed, move onto final test final test, if failed, move to step 8
   1. Final test
      1. 1 flight – Control of Steering, Altitude, Landing Gear, Flaps, Brakes, and Airspeed
8. Complete NASA-TLX about final test
9. Debriefing of overall study
10. Complete payment form

Results

Table 4 is a summary of the total results of the study showing the number of participants completing the training successfully and their results (pass/fail) on the final test by group. N represents the total number of participants in each group. The number of participants passing the training also indicates the number of people who were able to take the final test. This signifies the reason why the number of participants who took the final does not add up to the total number of participants in each group (N). This lowered the amount of total participants

(N) from 103 to 80. The total amount of each column is in the last row along with the average of the percent passing. More detailed results for the final test can be found in Appendix B.

Table 4: Results Summary of Pass/Fail for all groups (ranked by percent passing final). WTFF – Whole-Task Fixed Feedback, PTFF – Part-Task Fixed Feedback, WTEF – Whole-Task Elaborative Feedback, PTEF – Part-Task Elaborative Feedback, PTNF – Part-Task No Feedback, WTNF – Whole-Task No Feedback.

Group	N	# Pass Training	# Pass Final	Percent Pass Final
WTFF	18	17	8	47.06
PTFF	18	16	5	31.25
WTEF	18	12	3	25.00
PTEF	19	16	2	12.50
PTNF	14	11	1	9.09
WTNF	16	8	0	0.00
Total	103	80	19	20.82

Hypothesis 1: Part-task scenarios will result in more participants passing the final across all conditions

The Mann-Whitney *U* test was used to evaluate the hypothesis that part-task scenarios would produce a higher amount of participants passing the final test in comparison to the whole-task scenarios condition. The results of the test were not significant showing that there were no differences for task-type when comparing the amount of participants that passed or failed the final.

Hypothesis 2: Elaborative feedback will result in more participants passing the final across all conditions

The Kruskal-Wallis nonparametric comparison was used for evaluating the final test comparison of results (pass/fail) for each feedback type; elaborative feedback, fixed feedback, and no feedback. The comparison was used to evaluate the hypothesis that elaborative feedback would result in a larger amount of participants passing the training in comparison to the fixed feedback and no feedback conditions. Results showed a significant difference between the three feedback types (H=6.806, 2 d.f., *p*=.017) identifying that a difference did occur for the amount of participants passing the final between each groups. By performing a Mann-Whitney *U* test, results established that the difference was between fixed feedback and no feedback (z=-2.373, *p*=.009) showing that fixed feedback (n=13) had a larger amount of participant passing the final in comparison to no feedback (n=1). Percentages of passing the final test for each feedback type can be seen in Figure 6. Even though a significant difference was not seen between elaborative feedback and fixed feedback at an alpha level of .05, the graph shows that a larger percentage of participants passed the final in the fixed feedback condition when comparing it to the elaborative feedback condition.

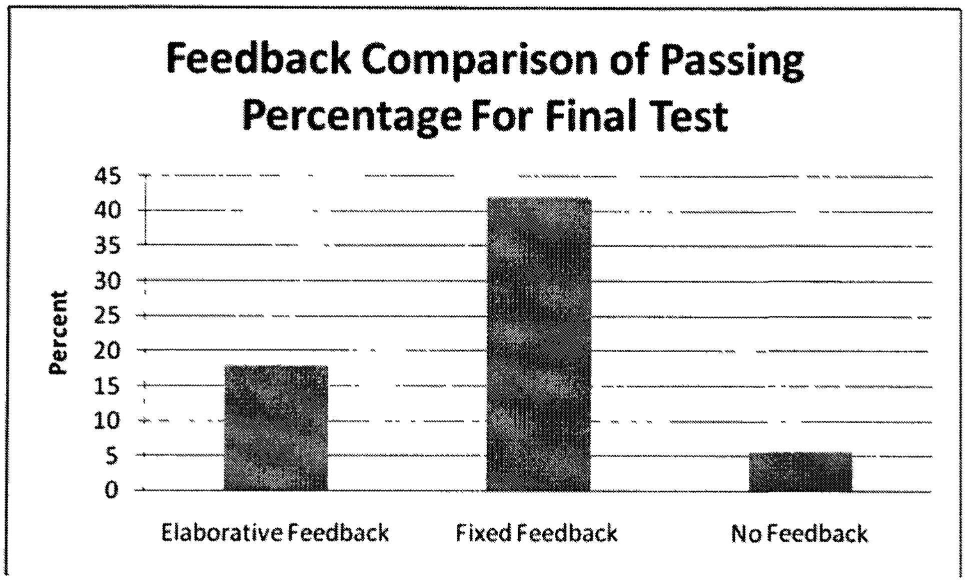

Fig. 6: Hypothesis 2: Percent passing the final test based on feedback type.

Hypothesis 3 Part-task scenarios with elaborative feedback will result in more participants passing the final in comparison to all other combined groups

The Kruskal-Wallis nonparametric comparison was conducted to test the hypothesis that participants with part-task scenarios and elaborative feedback would have a higher amount of participants passing the final in comparison to all other groups. The results of the test identified that there was no significant differences when comparing all groups during the final test for pass/fail. Even though no differences were found, the data shows that participants in the WTNF group (control) did not have one participant pass the final test, while almost half of the participants in WTFF passed the final as seen in Table 4.

Hypothesis 4 - No feedback will result in the lowest amount of participants passing the final (as this was a control group)

Please refer to the findings in the results section, *Hypothesis 2* and *Hypothesis 3.*

Hypothesis 5 Workload will follow the above hypotheses:

Task. A Mann-Whitney *U* test was performed against the different types of task scenario methods (part-task and whole-task) for all assessment criterions in the NASA-TLX workload assessment questionnaire. Workload assessments were based on how much mental and perceptual activity was required, how much physical activity was required, how much time pressure the participant felt based on the rate or pace of the task, how successful they were in accomplishing the goals of the task, how hard they had to work both mentally and physically, and how insecure vs. gratified they felt during the tasks. The test was used to evaluate the hypothesis that workload in all criterion would be lower for part-task in comparison to whole-task. Results indicate that there were no significant differences between the two groups for each individual criterion.

Feedback. A Kruskal-Wallis nonparametric comparison was once again used to compare all feedback groups for the different assessment criterion based on the NASA-TLX subjective workload assessment. This test was used to evaluate the hypothesis that elaborative feedback would result in the lowest amount of workload in comparison to fixed feedback and no feedback conditions across all criterion. When comparing the different feedback conditions for all assessment criterion, results found a significant difference for time pressure (H=7.901, 2 d.f., *p*=.019) showing a difference between the three feedback groups for time pressure. A Mann-Whitney *U* test was used to test the differences between groups for time pressure and results showed that there is a significant difference between elaborative and no feedback conditions (z=-2.690, *p*=.007). This finding indicates that participants with elaborative feedback (M=5.0833, SD=1.5) felt more time pressure than participants with no feedback (M=3.9, SD=1.668). The Kruskal-Wallis comparison between feedback conditions also resulted in a significant difference for feelings of being insecure vs. gratified (H=7.312, 2 d.f., p=.026) identifying a difference between feedback groups. A Mann-Whitney *U* test identified that there was a difference between elaborative and fixed feedback for feelings of being insecure or gratified (z=-2.624, *p*=.009) showing that participant with elaborative feedback (M=4.9722, SD=1.665) felt more gratified and less stressed during the task in comparison to the fixed feedback condition (M=3.8 SD=1.89) that felt more insecure and more stressed. The Mann-Whitney *U* also resulted in a difference between fixed feedback and no feedback (z=-1.848, *p*=.065) identifying that participants in the fixed feedback condition (M=3.8, SD=1.89) felt more insecure and stressed than participants in the no feedback conditions (M=4.7, SD=1.915). Table 5 displays all of the aforementioned results for workload comparison across feedback.

Table 5: Hypothesis 5 - Significant Workload Findings for Feedback. H/z is used to identify the Kruskal-Wallis statistics (H) or the Mann-Whitney U statistics (z).

Workload Measurement	**Comparison**	**H/z**	**d.f.**	**p**
Time Pressure	All Feedback Groups	H=7.901	2	0.019
	Elaborative vs. No Feedback	z=-2.690	N/A	0.007
Insecure vs. Gratified	All Feedback Groups	H=7.213	2	0.026
	Elaborative vs. Fixed Feedback	z=-2.624	N/A	0.009
	Fixed vs. No Feedback	z=-1.848	N/A	0.065

All Groups Comparison. Overall workload was analyzed using a Kruskal-Wallis nonparametric comparison for each criterion of workload assessment based on the NASA-TLX subjective workload questionnaire. The comparison was used to test the hypothesis that part-task scenarios with elaborative feedback would produce the lowest amount of workload in comparison to all other groups. When comparing all groups, results showed that there was a significant difference for feelings of goals (H=10.641, 5 d.f., *p*=.059) indicating that the groups differed in workload for how satisfied the participants felt for being able to accomplish their goals.

Hypothesis 6 - Training errors will follow the above hypothesis

Task. The two task-types (whole-task scenarios and part-task scenarios) were compared against each other to establish if there is a difference between the two for training purposes on the amount of passing or failing. The Mann-Whitney *U* test nonparametric comparison was conducted to test the hypothesis that participants with part-task scenarios would have a higher amount of participants passing training than whole-task. The results of the test were not significant showing that there was no difference in task-type for passing the training.

Feedback (Pass/Fail). A comparison was performed based on pass/fail for feedback types (elaborative, fixed, and no) across all conditions of task type during the training. The Kruskal-Wallis nonparametric comparison was conducted to evaluate the hypothesis that participants with elaborative feedback would have a higher ability to pass the training in comparison to fixed and no feedback conditions. Figure 7 displays the percentage of participants that passed the training for each level of feedback. The results indicated a significant difference between the three feedback types (H=6.823, 2 d.f., p=.017) showing that there is a difference between the feedback types for the ability to pass the training. By performing a Mann-Whitney *U* test, results showed that the difference was between fixed feedback with elaborative feedback, a significant difference was found (z=-1.780, p=.038) identifying that fixed feedback (n=31) had a higher amount of participants passing the training in comparison to elaborative feedback (n=28). A second comparison using the Mann-Whitney *U* test identified that there was a significant difference between fixed feedback and no feedback (z=-2.726, p=.003) showing that fixed feedback (n=31) had more participants passing the training than no feedback (n=18). Table 6 breaks down the two findings found by the Mann-Whitney *U* Tests. These results indicate that the fixed feedback conditions had a higher amount of participants passing the final in comparison to both elaborative and no feedback.

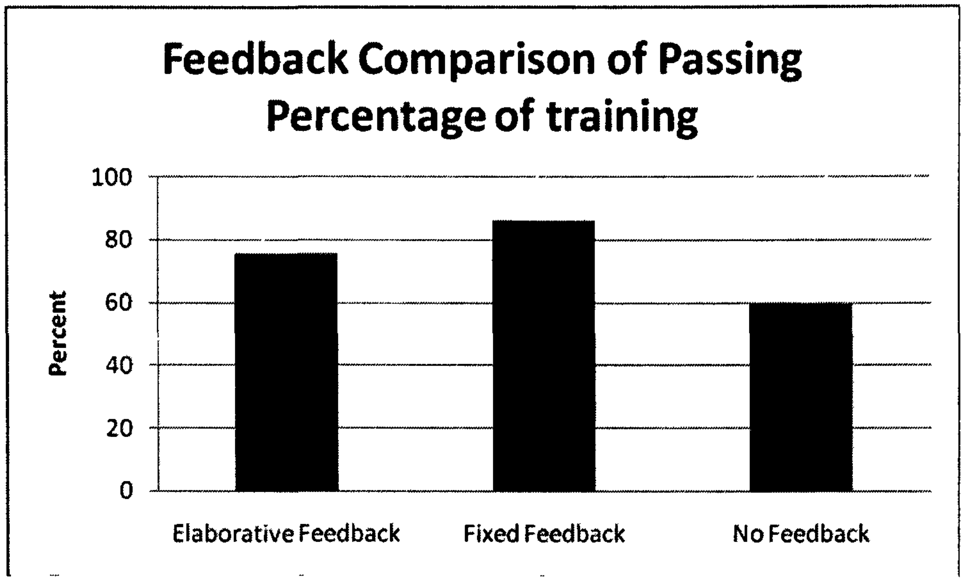

Fig. 7: Hypothesis 6 - Percent passing the training based on feedback type.

Table 6: Hypothesis 6 - Mann-Whitney U Test Significant Results for Feedback Conditions during training.

Comparison	Z	p
Fixed vs. Elaborative Feedback	-1.78	0.038
Fixed vs. No Feedback	-2.726	0.003

Comparison of All Groups (Pass/Fail). For the comparison of performance on training, non-parametric comparisons were made based on a Pass/Fail criterion. The percent passing rates of training and final test for all groups are show in Figure 8. The Kruskal-Wallis nonparametric comparison was conducted to test the hypothesis that participants with part-task scenarios and elaborative feedback would have a higher amount of participants passing the training than all other groups. The results of the test revealed that a significant difference was found between the different groups for the training segment of the study (H=12.593, 5 d.f., p=.013) identifying that not all groups performed the same during training.

A Mann-Whitney *U* test identified a difference between part-task elaborative feedback (PTEF) and whole-task no feedback (WTNF) (z=-2.141, *p*=.016) establishing that more participants in PTEF (n=16) passed the training in comparison to WTNF (n=8). Another Mann-Whitney *U* test revealed that a significant difference was found between part-task fixed feedback (PTFF) and WTNF (z=-2.353, *p*=.010) establishing that more participants in PTFF (n=16) successfully passed the training in comparison to WTNF (n=8). A third Mann-Whitney *U* test showed that a difference was found between whole-task elaborative feedback (WTEF) and whole-task fixed feedback (WTFF) (z=-2.076, *p*=.019) identifying that fewer participants in WTEF (n=12) passed the training in comparison to WTFF (n=16). A fourth Mann-Whitney *U* test showed a significant difference found was between WTFF and WTNF (z=-2.889, *p*=.002) establishing that more participants in WTFF (n=16) passed the training in comparison to WTNF (n=8). Table 7 displays the different findings from the Mann-Whitney *U* tests.

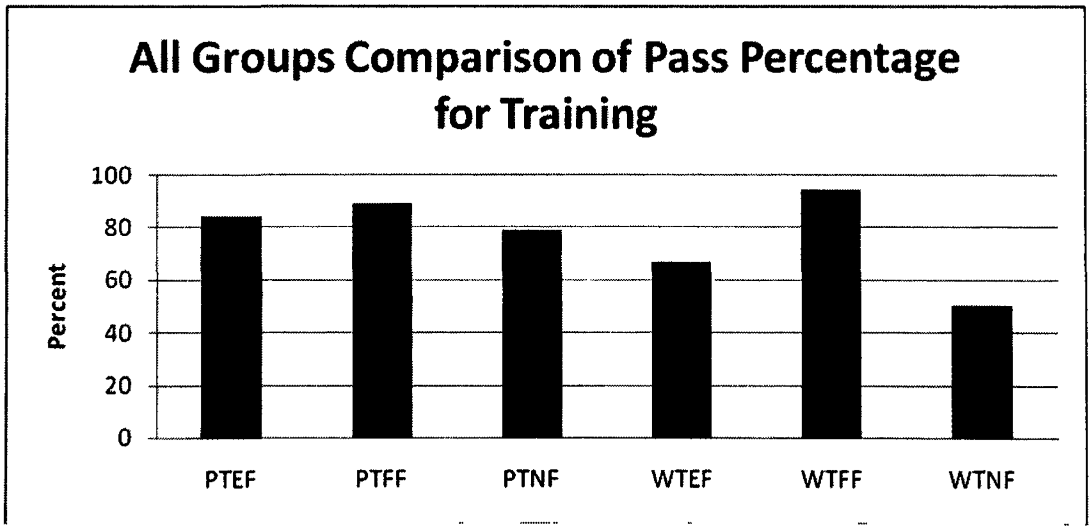

Fig. 8: Hypothesis 6 - Percentage of participants that passed the training. In WTNF condition, the percent passing rate was 0.

Table 7: Hypothesis 6 - Mann-Whitney U Test Significant Results across all conditions.

Comparison	**z**	**p**
PTEF vs. WTNF	-2.141	0.016
PTFF vs. WTNF	-2.353	0.01
WTFF vs. WTNF	-2.889	0.002
WTEF vs. WTFF	-2.076	0.019

Other Findings

Flight Simulation Experience on Final (Pass/Fail). A Kruskal-Wallis nonparametric comparison was used to test if there are any differences in performance for participants with different levels of experience of flight simulation video games. Figure 9 shows the percent passing the final test for participants based on their prior video game flight simulator experience. No participants with 0, 500-100, and 1000+ hours passed the final test. The lack of ability for participants to pass the training in the previously stated groups should be taken into consideration when comparing against the 0-100 hour group and 100-500 hour group since it caused the number of participants to be drastically smaller during the final test. No significant difference was found between the groups based on a pass/fail score.

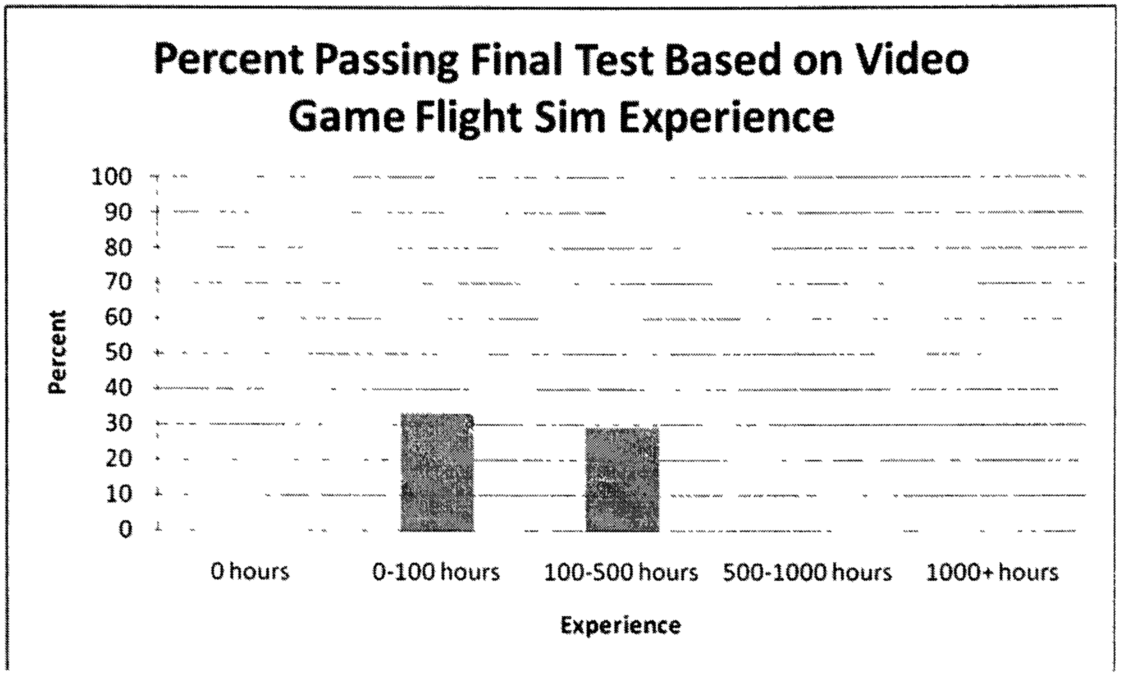

Fig. 9: Percent passing both training and final test based on video game flight simulation experience. Participants with 0, 500-100, and 1000+ hours did not have any participants that passed the final test.

Flight Simulation Experience on Training (Pass/Fail). In order to investigate if flight simulation experience would have an effect on performance during the training process based on a pass/fail result, a Kruskal-Wallis nonparametric comparison was used. Take note that there was a larger amount of participants with 0-100 hours in comparison to all other groups since students at Embry-Riddle are exposed to Flight Simulator video games during their private pilot training. Results indicate that no significant differences were found.

Group Differences between Scenarios. A repeated measures analysis of variance was conducted to test the differences between the three types of feedback on each scenario (scenarios 1 through 5) in the part-task condition based on overall scores by each participant. This test was conducted to evaluate if elaborative feedback would result in lower scores throughout the training scenarios. Results of this test indicate that there were no significant differences found in each of the scenarios based on the feedback type.

Figure 10 shows a graph of the average cumulative score committed by each feedback group in part-task based on the scenarios. As seen in the graph, the cumulative score/number of errors are similar for each group when looking at a single scenario, showing why no significant difference was found. On the other hand, when comparing the scenarios, scenario 5 (the addition of airspeed) results in a much larger amount of overall errors in comparison to the other scenarios.

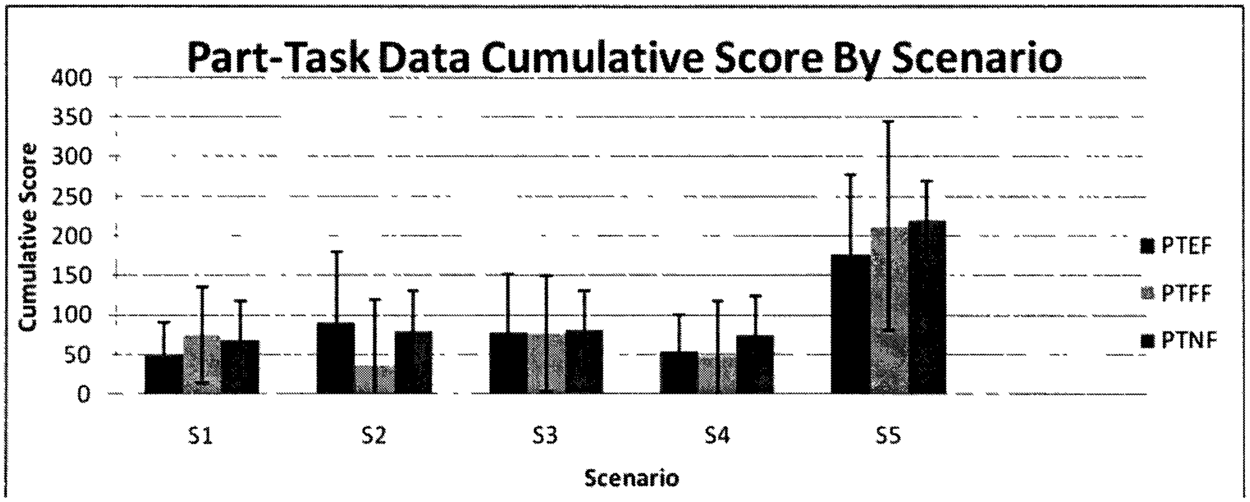

Fig. 10: Cumulative Scores for each feedback group throughout the part-task training scenarios. PTEF, PTFF, and PTNF can be read by following from left to right for each scenario (S1-S5). Error bars signify standard error for each group.

Final Score Comparison for All Participants Taking the Final. Due to a large number of pilots not passing the final, the last analysis focus on all of the pilots who were eligible to take the final. In Figure 11, all the scores on the final for each group consisting of both passing and failing scores are shown. The graph shows the average score for each group, +/- the standard deviation of the mean. The lower scores are once again associated with the best performance as in less errors occurred. The data consisting of the final cumulative scores by each participant taking the final test were analyzed using the Kruskal-Wallis test which again showed no differences between each group based on their score on the final test.

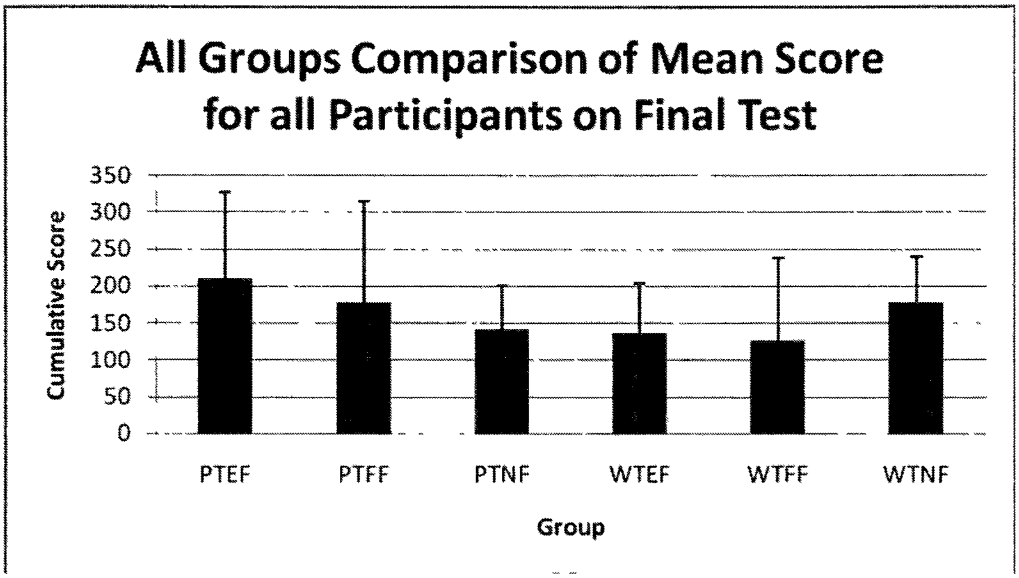

Fig. 11: All Groups Comparison of Mean Score for all Participants on Final Test. Error bars indicated standard deviation of each group.

Discussion

Experimental Findings

This study investigated the effects of task-training techniques and feedback on the ability to reach and maintain a specified standard during a simulated Boeing 737-800 landing into Seattle-Tacoma International Airport. Furthermore, it examined if the task-scenario can be combined with a specific type of feedback to increase learning comprehension while lessening the required time to be able to complete a specific task. This study was the first time this complex stick and rudder task was used for research due to the newly developed Microsoft ESP flight simulation program and the scenarios built by Boeing Phantom Works.

While 103 participants were used in total, only 80 participants reached the final test and could be used for evaluation for final test scores (pass/fail). Out of these 80 participants, only 19 passed the final test identifying a dramatic failure rate. This was an ambitious study that looked into expediting the amount of time needed to train a student to perform a complex task such as an ILS approach. Results shows that not enough training was given to the participant before passing them on and more training is needed to bring all participants up to a required skill level that will allow them to successfully pass the ILS task with no feedback. Based on the time it takes to train an ILS approach in most flight schools, approximately three hours is just not enough time to expect students to learn how to execute these tasks at such a high level of performance. One exposure of training is not enough to allow for participants to successfully land the 737 aircraft without any feedback.

In total, the study was concerned about six main objectives. The first main objective was to investigate if a specific task-scenario technique would result in a higher amount of participants being able to pass the final test. No difference in the ability to pass the final test was seen for task type. Previous investigations of part-task versus whole-task are inconclusive for explanations on which causes better overall performance during a complex task such as landing a simulated 737 and this study concludes that there are no differences between the two for this task. Reasons for no differences occurring are a result of the dramatic failure rate exhibited during the final test in all groups. As explained above, the final test was too difficult for the participants based on their skill level after the training. If training was completed over a series of exposures instead of only one 3-4 hour period, results may differ. A better structure is needed to expand this training into a prolonged period of exposures to see differences between the two training styles.

The second main objective was to evaluate if feedback styles caused a difference in the ability to pass the final test. Fixed feedback showed a significantly higher amount of participants passing the final test than no feedback showing that over the control group, a difference did occur. Based on a $p<.05$ alpha value, fixed feedback did not show a difference over elaborative feedback, but Figure 6 provides a clear example that almost half of the participants that reached the final test in fixed feedback conditions passed the final while elaborative feedback groups only resulted in approximately 17 percent of participants passing. The low amount of participants and unequal N's of groups for participants that took the final could have had an effect on the ability to see a difference, but as the graph show, fixed feedback shows the greatest improvement in ability to pass the final test.

These findings are once again against the original hypothesis that elaborative feedback would have increased the ability for the participant to learn how to complete the task, but in hindsight, participants may have been overwhelmed by the amount of information presented through the auditory cues. The 737 landing scenario was a very complex task with many warnings capable of occurring. If multiple errors were committed at one time, a multitude of error warnings may have occurred, overloading the participants processing abilities. Elaborative cueing may have provided too much information that was not necessary for the participant to understand, locate, and fix an error, while fixed feedback gave limited information to "check" an error without becoming a distraction. This supports Morrison, Ross, Gopalakrishnan, and Casey's (1995) statement that elaborative feedback can be too costly to develop and does not provide a great enough benefit to be created. The detailed information found in elaborative feedback did not increase performance as previously hypothesized when provided to the participant as seen when comparing to other types of feedback with simple cues as seen in this study with fixed feedback. In addition, private pilots from Embry-Riddle are some of the best trained pilots in their skill level based on the highly structured training that they receive. A simple check may have been enough for these pilots to identify the problem and quickly fix it without removing their attention from the main task of landing the aircraft. On the other hand, no feedback did not give the pilot enough information, resulting in a high amount of students failing the final test.

Also, the presentation of the auditory prompts (feedback) was given using the Microsoft text-to-speech Mary voice. The Microsoft text-to-speech voice was used to allow the participant to distinguish the difference between co-pilot/embedded instructor commands (human recorded voice) and error/feedback cues (text-to-speech voice). This presentation method may have caused confusion on what the elaborative cues were trying to explain during the third level of cues (if an error was committed for a third time consecutively). At the third level of cues, a small sentence was given to the participant to explain how to fix the error. The length and voice may have lead to participants not being able to fully understand what was being told to them. The first level cue (at the initial start of an error) that was presented to participants in elaborative feedback was the same as those presented to participants in fixed feedback. Short phrases such as "Check Heading" may have been clear enough for participants to understand the statement that was given to them, but anything longer may have been unclear due to the simulated text-to-speech voice.

The third finding of this study looked at the interaction between the task scenarios and the type of feedback that a participant was presented against the ability the pass the final test.

Based on the previously mentioned findings, no interaction was found between the two due to the low amount of participants able to pass the final test. The flooring effect exhibited in this study diminished the ability to find a difference between the groups because of the need for a longer amount of training on a complex task such as this one. If more time was given to the participants to train, the ability to see a difference (if one exists) would greatly increase. Compared to all other groups, whole-task fixed feedback (WTFF) provided the largest amount of participants passing the final with approximately half, while whole-task no feedback (WTNF) resulted in the lowest amount of participants reaching (n=8) and passing the final (n=0). When looking at all the groups together, WTFF had a higher amount of participants passing the training in comparison to whole-task elaborative feedback (WTEF) which is opposite of what was originally hypothesized. Once again, it was believed that elaborative feedback would result in more passing scores based on previous literature. WTNF was a control group placed into the study to see if exposure to the final test at the maximum amount of exposures (20 exposures) would result in participants being able to pass without any means of feedback or training style.

With these statements about WTNF, it now leads into the fourth objective which is that no feedback conditions would result in the lowest performance (amount of participants passing the final) in comparison to all other groups. During the training, the whole-task no feedback group which was used as a control was outperformed by almost all other groups. This helps support the notion that exposure did not influence results for passing the final test since participants that were given up to 20 attempts on the same scenario as the final test (WTNF) still scored well above the passing 100 points criteria throughout the training (≥100 points = failure). The participants that also reached the final in this condition (WTNF) did not pass the final test which is counter-intuitive to what one would think since they previously passed the same flight.

Whole-task with no feedback provided no guidance to what the participant was doing incorrectly or how to fix an issue. The pilots had to rely on their previous knowledge of how to fly a standard aircraft using instruments even though they had little to no prior experience with instrument flight. It was expected that a low amount of participants would make it to the final due to how difficult the task was without any feedback which was demonstrated in this study.

The fifth objective focused on the differences exhibited by task scenario styles and feedback methods for different types of workload measurements based on the NASA-TLX subjective workload assessment questionnaire. The NASA-TLX subjective workload assessment was used due to it being available in paper and pencil, its previous background of being used as a beneficial research tool, and how relative the questions are to the task at hand. The test allows for one to understand specific questions on not only physical workload, but mental workload including time pressure, successful feeling of goals, and how irritated/stress/insecure versus gratified one feels during the entire task. Based on previous research, whole-task exhibits a larger amount of workload in comparison to part-task training since it forces students to adapt to a highly complex situation from the initial start of training (Gopher, Weil, & Siegel, 1989). Workload in the beginning of whole-task is extremely high in comparison to part-task, but is the same as exhibited by part-task towards the end when part-task reaches the same level of complexity as whole-task. With these statements, it was hypothesized that overall workload in the whole-task scenarios would be highest for all criterion in the NASA-TLX workload assessment questionnaire in comparison to part-task. Results of this study indicate that there were no differences in all measures for each criterion thus disproving the initial hypothesis. Participants may have seen that the overall workload throughout the study was the same due to the increasing complexity caused by part-task compared to the initial high workload that decreased as more exposures occurred in whole-task.

When comparing the different type of feedback methods for workload, it was originally hypothesized that elaborative feedback would have a lowest amount of overall workload throughout the study due to it guiding participants in a step by step process to identify and fix an error based on their own individual needs. This study showed that elaborative feedback and no feedback participants felt more gratified and less stressed during the task in comparison to fixed feedback. When judging this against the previously stated results of fixed feedback causing the highest amount of participants passing the training, this amount of workload may have helped the participant. The extra stress could have caused the participant to want to work harder to accomplish the goal and diminish the feeling of irritation created throughout the training. The study also showed that elaborative feedback also had a higher time pressure throughout the entire study than the no feedback condition. The rate of events occurring may have seemed a lot faster due to the complexity of the information being given to the participant every time an error occurred. This could have lead to an added stress that counteracted with the ability to successfully reach and maintain standards.

The sixth objective was looking at the differences exhibited during training for the ability to pass onto the final test. When comparing the different training scenarios (part-task and whole-task) no differences were found once again. As stated previously, in order to see a difference, a larger amount of time/exposures through these training styles are needed. The amount of participants that accomplished the task of completing training depreciated by 23 participants out of a total of 103. Participants showed likelihood to pass through the training in both conditions, but when they flew in the final test, only 19 passed. The training for both task conditions needs to be expanded into a longer amount of exposures in order to prepare the students for the final.

When comparing the different feedback conditions for the ability to pass the training, fixed feedback lead to more participants reaching the final in comparison to elaborative feedback and no feedback conditions. Previously explained, the amount of information given to the participant was limited, only providing them with information on the error that was occurring. This may have been enough for them to understand what was happening without needing the extra knowledge of how to fix it since they are already a well-skilled pilot (based on having a private pilot's license). Too much information may have confused the pilot and pulled their attention away from the complex task, while no feedback did not provide enough information to benefit learning.

Lastly, when comparing all the groups, whole-task scenarios with no feedback had fewer amounts of participants reaching the final than all groups, specifically against part-task elaborative, part-task fixed feedback, and whole-task fixed feedback. Once again, WTNF was a control group and it was expected that a small number of participants would pass the training due the participants being given no information on what errors were being committed (reason for failure). This comparison also showed a difference between whole-task fixed feedback (WTFF) and whole-task elaborative feedback (WTEF) in that WTFF lead to more participants passing the training than whole-task elaborative feedback. Contrary to the original hypothesis that when comparing the same task type and different feedback, elaborative feedback would have provided the most beneficial information for mastering the task, fixed feedback provided the participants with the small amount of information needed which resulted in the most amount of participants passing the training.

Other data collected during this study investigated if previous video game flight simulation experience had an effect on the ability to pass the training and the final test. The video game experience shown by participants was not an independent variable in this study and this information was collected through the experience questionnaire administered prior to beginning the simulation. Participants had this experience on their own time prior to being in this study. This study showed those participants with 0 hours, 0-100 hours, 100-500 hours, 500-1000 hours, and 1000 + hours of video game flight simulation experience did not exhibit any differences in ability to pass the training or the final test. This is counter-intuitive to natural thought that someone who understands how basic Microsoft Flight Simulator (MFS) works would have an advantage over others in a task on MFS. The structure of the training along with the familiarization period may have allowed participants with low video game flight simulation experience to be able to be at the level understanding of the controls and environment as those with higher levels of experience. This shows that the training techniques (both part-task and whole-task) may be beneficial to all levels of experienced gamers that are utilizing a desktop simulation video game as a training aid in this task. To note, most participants in this study had 0-100 hours of flight simulation experience since Embry-Riddle Aeronautical University exposes flight students to MFS during their flight training curriculum as a supplemental training tool. Having more participants with a variety of MFS and other video game flight simulation experience may be beneficial in looking more into why there is no difference seen.

Another set of data that was analyzed compared the total amount of errors/cumulative score across each scenario in part-task, which did not show any effect over the different scenarios when comparing the three different types of feedback. All feedback conditions provided the same amount of guidance to participants in each scenario across the different scenarios. One of the reasons that this may have occurred is due to the fact that participants crashes could not be calculated into the study since the number of errors did not reflect the total flight. If crashes were calculated into the overall cumulative score for each group, scores may change, showing a difference between the feedback types.

Figure 10 shows the different scores between groups for each scenario, allowing one to notice a large increase in cumulative score for scenario 5. This is due to the addition of airspeed into the training regimen. Shebilske et al. (1999) stated that one should automate the most difficult element of a task in the hierarchical part-task scenarios. As time goes on, the most difficult tasks should stay automated until the less difficult tasks are mastered. This study shows that the most difficult task, airspeed, is being shown too late and may have had an effect on the cumulative score and the amount of participants passing the final test in part-task conditions. Participants may have moved onto the final test even though they never mastered the airspeed (most difficult) portion of the task in the landing scenario due to the logic placed into the training. The logic of only 4 attempts per scenario for part-task training was used to allow for the same amount of attempts for all participants regardless of task-training type (20 attempts max). If participants were given this task at an earlier time, they may have been able to master this skill and score a lower cumulative score (less total errors) during the final test. Also, the amount of time given to master this task may have affected the total score on the final test as well as scenario 5. If the amount of time/exposures are increased, participants may be able to master the task and eventually pass the final test. Previous studies such as Gopher, Weil, and Siegel (1989) have stated that it is difficult to isolate and structure part-task scenarios and with this knowledge, this task can be restructured for future research.

The last set of data looked at each individual cumulative score on the final test for all participants eligible to reach the final test. A Kruskal-Wallis showed no significant difference once again for the comparison of all the groups, demonstrating that the original Pass/Fail criterion also gave the correct (same) findings as previously explained. More time exposures and a larger/equal number of participants in each group may allow for differences between the groups to occur. Again, the data collected during this study once again supports that the training did not cause a large enough difference between each of the conditions to discern an effect. This is most likely due to a longer amount of time in training needed since learning to land a Boeing 737-800 aircraft and an ILS approach is too difficult to teach in a 3-4 hour period.

Future Research

This study's utilization of Microsoft ESP as an experimental tool for research has not been performed prior to this study due it being the most recent addition to the Microsoft Flight Simulator series. Future research can make use of this tool to develop and conduct a whole host of different types of research since it is an open source program that allows any program to run as an assessment in the background. This study's investigation of different task types should be further investigated on the placement of the different scenarios in part-task to see what will create the greatest ability reach and maintain standards throughout the training and into the final test. This can be performed by changing the rotation of the different scenarios such as placing scenario 5 (airspeed) into an earlier scenario instead of introducing it at the end. This may unveil new ideas for the structuring of hierarchical part-task training contrary to previous research. More testing should also look at the number of scenarios or exposures used. Due to the complex nature of this task, more scenarios and exposures may be useful to help train the participant to a higher level of ability since their ability after training was not high enough to successfully complete the difficult final test.

Different cueing techniques should be investigated to see if the reason for elaborative cueing not performing as well as originally hypothesized is due to it being an overload of information or if it is based on the presentation style. Since development of elaborative cueing is difficult as stated in previous research, studies should be performed testing the amount of information in elaborative cueing as well (i.e. how much detailed information is given to the student?). This would allow the development to be created to the exact amount needed to decrease overall scores and increase the ability to reach and maintain standards. Also, tests should be conducted comparing text-to-speech versus human voice presentation along with studies to see if visual cues may increase the ability to process the high amount of information being presented. The presentation style can cause a large effect on the ability to process information and with highly complex commands that are stated in elaborative feedback, other mediums may result in better learning during a complex task.

Conclusion

The environment of pilot training is altering due to a need for an increase in the amount of instructors and pilots, triggering new styles of training such as adaptive training scenarios to be produced that will lessen the amount of time an instructor needs to be with a student. Through the use of a performance assessment engine that captures real-time statistics, a student's performance can be quickly analyzed to understand their errors and achievements. Adaptation in a simulation that is based on performance rather than time leads to many new challenges including the type of cueing and training method that a student would need to help improve their overall performance. The nature of different styles of feedback/cueing and the method to train a student in an adaptive situation was studied and showed that while the training methods caused no difference in ability to pass the training and reach the final test, the type of feedback produced a difference in the ability to reach and maintain a pre-determined standard. The study that was conducted did not allow enough training to make a full comparison of the task-types (part vs. whole), but with more exposures, a better comparison can be made in the future. With the findings that were found, one could employ fixed feedback into a desktop training simulation allowing for many advantages to possibly occur. It can help to create training programs that could decrease the amount of cost since there is less time a student will need to be with an instructor in a high fidelity aircraft simulator and increase the ease of training since it can be performed on an average desktop computer. Tools such as training simulations that are developed with the correct training style and feedback can be helpful to circumvent the unnecessary practice of procedures and techniques that have already been learned by a student.

References

Alessi, S. M. (1988). Fidelity in the Design of Instructional Simulations. *Journal of Computer-Based Instruction, 15*(2), 40-47.

Corbalan, G., Liesbeth, K., & Merrienboer, J. (2008). Dynamic task selection; Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction*, 1-13.

Dennis, K.A, & Harris, D. (1998). Computer-Based Simulation as an Adjunct to Ab Initio Flight Training. *The International Journal of Aviation Psychology, 8*(3), 261-276.

Ende, J. (1983). Feedback in Clinical Medical Education. *Journal of the American Medical Association, 250*(6), 777-781.

Ennels, A. (2002). The "Wright stuff" pilot training at America's first civilian flying school. *Air Power History, 49(*4), 22-32.

Fabiani, M., Buckley, J., Gratton, G., Coles, M. G., & Donchin, E. (1989). The training of complex task performance. *Acta Psychologica, 71*, 259–299.

Gagne, R.M. (1954). Training Devices and Simulators: Some Research Issues. *The American Psychologist, 9,* 95-107.

Goettl, B. P. & Shute, V. J. (1996). Analysis of part-task training using the backward-transfer technique. Journal *of Experimental Psychology: Applied, 2*(3), 227-249.

Gopher, D. (1996). Attention control: Explorations of the work of an executive controller. *Cognitive Brain Research, 5*, 23–38.

Gopher, D. Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors, 36*, 387–405.

Gopher, D. Weil, M., & Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica, 71*, 147–177.

Hart, S.G. (2008). NASA-Task Load Index (NASA-TLX); 20 Years Later. NASA-Ames Research Center.

Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Human Mental Workload*, 239-250.

Hummel, H., & Nadolski, R. (2002). Cueing for Schema Construction: Designing Problem-Solving Multimedia Practicals. *Contemporary Educational Psychology, 27*, 229-249.

Hummel, H., Paas, F., & Koper, E.J.R. (2006). Timing of cueing in complex problem-solving tasks: Learner versus system control. *Computers in Human Behavior, 22*, 191-205.

Kawashima, R., et. Al. (2000). The Effect of Verbal Feedback on Motor Learning – A PET Study. *NeuroImage, 12*, 698-705.

Kelly, L. (1970). *The pilot maker.* New York, NY: Grosset & Dunlap.

Koonce, J.M., & Bramble, W.J. (1998). Personal Computer-Based Flight Training Devices. The *International Journal of Aviation Psychology, 8*(3), 277-292.

Kramer, A. F., Larish, J. F., & Strayer, D. L. (1995). Training for attentional control in dual task settings: a comparison of young and old adults. *Journal of Experimental Psychology:Applied, 1*(1), 50-76.

Kulhavy, R., & Wagner, W. (1993). Feedback in Programmed Instruction: Historical Context And Implications for Practice. *Interactive Instruction and Feedback*, 3-20.

Kulhavy, R., White, M., Topp, B., Chan, A., & Adams, J. (1985). Feedback Complexity and Corrective Efficiency. *Contemporary Education Psychology, 10*(3), 285-291.

Liu, D., Macchiarella, N. D., & Vincenzi, D. A. (2008). *Human Factors in Simulation Training.* Orlando, FL:Parntership 1.

Longridge, T., Burki-Cohen, J., Go, T.H., & Kendrea, A.J. (2001). Simulator Fidelity Considerations for Training and Evaluation of Today's Airline Pilots. *Proceedings of the 11th International Symposium on Aviation Psychology (pp. 1-7).* Columbus, OH: The Ohio State University Press.

Lorell, M. A. (2003). *The U.S. Combat Aircraft Industry, 1909-2000: structure, competition, innovation.* Santa Monica, CA. RAND National Defense Research Institute.

Magerko, B., Wray, R., Holt, L., & Stensrud, B. (2005). Customizing Interactive Training

Through Individualized Content and Increased Engagement. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).*

Morrison, G.R., Ross, S.T., Gopalakrishnan, M., & Casey, J (1995). The Effects of Feedback and Incentives on Achievement in Computer-Based Instruction. *Contemporary Educational Psychology, 20*, 32-50.

Narciss, S., & Huth, K. (2002). How to design informative tutoring feedback for multi-media learning. *Instructional design for multimedia learning*, 1-16.

Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction, 16*, 310-322.

Ortiz, G., Kopp, D., & Willenbucher, T. (1995). Instrument training using a computer-based simulation device at Lufthansa Airline. In R.S. Jensen & L.A. Rakovan (Eds.), *Proceedings of the Eight International Symposium on Aviation Psychology* (pp. 1047-1050). Columbus: The Ohio State University.

Oser, R. L., Cannon-Bowers, J.A., Salas, E. & Dwyer, D. J. (1999). Enhancing human performance in technology-rich environments: guidelines to scenario-based training. *Human/Technology Interaction in Complex Systems, 9,* 177-202.

Ramachandran, S., Cramer, M., Harville, L., & Ashworth, A. (2003). Developing an Adaptive

ADL Solution for Training Medical Teams. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).*

Smode, A.F., Hall, E.R., & Meyer, D.E. (1966). *An Assessment of research relevant to pilot Training* (*Vol 11.)* (Tech Rep No AMRL-TR-66-196). Wright Patterson Air Force Base, OH; Aerospace Medical Research Laboratory.

Shebilske, W., Goettl, B., Regian, J. (1999). Executive Control and Automatic Processes as Complex Skills Develop in Laboratory and Applied Settings. *Attention and Performance XVII*, 401-432

Waldersee, R., & Luthans, F. (1994). The impact of positive and corrective feedback on customer service performance. *Journal of Organizational Behavior, 15*, 83-95.

Werrell, K.P. (2005). Those Were the Days: Flying Safety during the Transition to Jets, 1944-53. *Air Power History, 52*(4), 38-54.

Wightman, D. C., & Sistrunk, F. (1987). Part-task training strategies in simulated carrier landing final-approach training. *Human Factors, 29*, 245–2.

Xiangton, Q., Bard, J., & Gang, Y. (2004). Class Scheduling for Pilot Training. *Operations Research, 52*(1), 148-163.

APPENDIX A

Consent Form

CONSENT FORM

Embry-Riddle Aeronautical University

I consent to participating in the research project entitled:

The Effects of Cueing during Simulated Flight Training

The principle investigator of this study is: Jon French, a professor within the Human Factors Department at Embry-Riddle Aeronautical University.

The individual above, or their research assistants, have explained the purpose of the study, the procedures to be followed, and the expected duration of my participations. Possible benefits of the study have been described, as have alternative procedures, if such procedures are applicable and available.

Basically the study will assess different training methodologies for flying a Boeing 737 aircraft using Microsoft Flight Simulator (ESP). Additionally, I consent to completing questionnaires providing some detail about my subjective experiences during the training. I will be expected to learn the skills needed and the computer will score my performance throughout approximately 15 flight scenarios. I will do the best I can to learn to fly the 737 and will not interfere with or distract the other participants in the flight simulation environment.

I will arrive at the testing site in the Human Factors Laboratory (LB374) at the assigned time agreed upon by the experiment, Dr. French, and myself. I will pledge approximately 4 hours until all documents and the training scenario is fully completed. I understand that all my records will be safeguarded and kept anonymous by Dr. French.

I am being paid $100 for my participation in the one session which could last from 3-5 hours. I understand the study is designed to benefit the study of training in low level flight simulators. I have been assured that my data and my participation in the study will be kept anonymous and will only be reported in the aggregate, as a group average and not as an individual score.

I acknowledge that I have had the opportunity to obtain additional information regarding the study and that any questions I have raised have been answered to my full satisfaction. Furthermore, I understand that I am free to withdraw consent at any time and to discontinue participation in the study without prejudice to me. However, if I choose to end my participation in the study before I have completed all the training conditions, I will forfeit any reimbursement and I will not be paid for my participation.

I agree not to disclose any information about the study to anyone who might possibly be a participant in the study. My experiences might not be what they experience as participants and my disclosure might bias or prepare them unfairly and incorrectly and would affect the outcome.

Finally, I acknowledge that I have read and fully understand the consent form. I sign it freely and voluntarily. A copy has been given to me.

Date: ________________________

Name (please print):____________________________

(*Participant*)

Signed:_____________________________________

(*Participant*)

Signed:______________________________________

(*Researcher/Assistant*)

Experience Questionnaire

Participant #:_________ Date: ____________

Experience Questionnaire

Thank you for participating in this study. This form asks questions about your piloting and computer experience. We will use this information to determine any factors that might explain your proficiency in flight and simulation. Please answer the following questions:

Name: ______________________________
Age:__________________
Sex: Male Female
Phone Number: ________________
Email: _______________________
Flight Licenses and certificates achieved:
(List each, the approximate date achieved, and years held, and if current. Use additional paper if more room is needed.)

License/Certificate	Date Achieved	Years Held	Current? Yes/No

Total logged Hours (in all aircraft): ________________
Last 12 months:_______________ Last 90 Days: ________________

Please list any recurrency training you've received and the date.

____

Do you have any experience with/in a 737 (Including Flight Sim)?
Yes No

If Yes, Please indicate what experience and how much time (hours):

____

Do you have experience playing flight simulator video games?
Yes No

If Yes, Please specify which game and the amount of experience (circle one):
0 - 100 hours 100 - 500 hours 500 - 1000 hours 1000 hours +

Instructions Example

Instructions to Subjects: Whole-Task Hierarchical Cues

Thank you for participating in our research study. The purpose of this research is to determine the most effective ways to train pilots in individualized computer-based instruction. In this study, we will be teaching you to land the Boeing 737-800 aircraft at Seattle SEA-TAC airport. There are a number of tasks that must be accomplished in order to perform this task successfully. The training is designed to take you through all of these steps and to teach you proficiency at this task. These instructions describe in detail what will happen during the training session.

Informed Consent and Background Questionnaire
Each participant will be asked to sign an informed consent form, and to complete a background questionnaire which asks for details regarding previous flight experience. Please fill this out as accurately as possible.

How to Fly the 737-800 Scenario
In the scenario overview, we will provide auditory familiarization for the basic controls and tasks required to fly and land the 737-800. Prior to beginning each mission scenario, there is a Mission Briefing window that provides you with some basic information. Read this briefing and then click in the upper right hand corner to close this window. When ready to proceed, hit *p* on the keyboard to begin the flight. The simulated instructor will tell you how to perform certain tasks, and then ask you to perform them. Once you have successfully completed that task, instructions for an additional task will be given. Please refer to the labeled drawings of the main instruments at any point during this orientation for clarification. IT IS IMPORTANT TO RE-SET ALL FLIGHT CONTROLS TO THE "UP" POSITION AT THE END OF EACH SCENARIO.

Landing Scenario Overview
Once you have successfully completed the basic flight orientation, you will be presented with a landing scenario walk-through. In this scenario, the instructor will walk you through the landing scenario into SEA-TAC airport. He will describe where to turn, when to intercept the glide slope, target airspeed, etc. In this scenario, the instructor will fly the approach and landing. You are simply observing this flight and listening to the instructor.

Pre-Test
Once you have completed the how-to-fly scenario and the landing overview, we will ask you to complete the pre-test flight. During this flight, all flight controls will be manual (no autopilot), and there will be no performance cues. You will receive basic audio instructions from the co-pilot, but you have total responsibility for the flight controls and landing. This flight is meant to serve as a baseline for performance. You will fly this identical flight again at the completion of the training scenarios. Perform to the best of your ability.

Training Scenarios

Following the pre-test, you will perform a series of training scenarios. For each training scenario, you will be controlling all aspects of the aircraft, identical to the pre-test. Your performance will be evaluated against targeted measures. If your performance exceeds a specified criterion, you will hear an auditory warning. This warning should serve to help you correct performance. If you continue to exceed performance tolerances, you will hear a more detailed audio warning. Cues become progressively more detailed the longer you exceed acceptable performance limits. Try to correct performance each time that you receive a warning. If you complete the given scenario with a high enough score, you will pass the training scenario and proceed to the final evaluation. If you fail a given training scenario, you will hear a cue stating that you need a little more work on the scenario. It will then restart the training scenario. In some instances, the scenario will end mid-flight if you have over-stressed the aircraft or exceeded the flight boundaries. The training scenarios will continue to repeat until you eventually pass, or until you have 20 unsuccessful attempts.

At the end of certain scenarios, you will see a PowerPoint Hints file pop up, or appear as an orange indicator on the menu bar at the bottom of the screen. Open this file and review the material prior to continuing with practice. This file provides pointers, strategies and tips for mission success. A Mission Briefing window will appear onscreen prior to each training scenario. It will describe your areas of responsibility for the upcoming scenario. As in the mission overview, read through the Mission Briefing and then close the window prior to initiating the training scenario. Hit *p* on the keyboard to begin each scenario. AGAIN, REMEMBER TO RE-SET THE FLIGHT CONTROLS PRIOR TO EACH SCENARIO.

After each training scenario, we will ask you to complete a subjective workload assessment questionnaire. This questionnaire will assess how difficult you feel that each scenario is, and how much effort you exerted to perform the required tasks. Specific details of this assessment will be explained prior to the first assessment.

Post-Test Scenario

Once you complete all levels of training, you will be asked to fly a post-training scenario. This scenario will be identical to the pre-test completed earlier. This test is a measure of how much you learned during the training, so please try your best. At the end of the post-test, you should hear a message to "inform instructor that training is complete." At this point, please notify the instructor on duty that you have completed the testing and wait further instructions.

Thanks for your participation.

NASA TLX

Place an X on the line in the position which best describes your evaluation for the flight you just made.

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

|--------|--------|--------|--------|--------|--------|--------| N/A
Low High

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

|--------|--------|--------|--------|--------|--------|--------| N/A
Low High

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

|--------|--------|--------|--------|--------|--------|--------| N/A
Low High

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

|--------|--------|--------|--------|--------|--------|--------| N/A
Failure Perfect

How hard did you have to work (mentally and physically) to accomplish your level of performance?

|--------|--------|--------|--------|--------|--------|--------| N/A
Low High

How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and competent did you feel during the task?

|--------|--------|--------|--------|--------|--------|--------| N/A
Low High

Payment Form

Participant #:______________
Date: ____________________

Participant/Rater Payment Sheet
Thank You for your help

Please Provide a GOOD email, phone number and address for you where we can reach you

NAME____

ADDRESS____

PHONE____

EMAIL____

Are you currently employed by the University? YES NO EXPLAIN

Are you currently a student at ERAU? YES NO

Please indicate your student ID # ___________________________

In order to pay you the right amount, please indicate your approximate hourly salary from the university

---

GOOD FLYING!

Stop by in a few months if you're interested in how the study turned out.

Appendix B

Detailed Demographic Table

Part Task Elaborative Feedback

Participant Number	Gender	Age	Flight Hours: Total	Flight Hours: Last 12 Months	Flight Hours: Last 90 Days	Ratings	Flight Sim Experience	737 Sim Experience
PTEF1-B1	M	20	105	15	0	Private Pilot	0-100 hours	No
PTEF2-B2	M	21	90	15	0	Private Pilot	0-100 hours	Yes - Flight Sim
PTEF3-B3	M	19	70	60	10	Private Pilot	0-100 hours	No
PTEF4-B1	M	18	80	80	15	Private Pilot	100-500 hours	Yes - FS
PTEF5-B2	M	18	97	30	10	Private Pilot	0-100 hours	No
PTEF6-B3	M	19	110	0	0	Private Pilot	500-1000 hours	Yes - Flight Sim
PTEF7-B3	M	23	58.9	48	0	Private Pilot	0-100 hours	No
PTEF8-B2	M	19	122	100	15	Private Pilot	0-100 hours	No
PTEF9-B1	F	18	70	70	0	Private Pilot	0-100 hours	No
PTEF10-B2	F	20	115	35	15	Private Pilot	0 hours	No
PTEF11-B1	M	23	120	120	10	Private Pilot	0 hours	No
PTEF12-B4	M	20	58	58	3	Private Pilot	1000+ hours	Yes - Flight Sim/X plane
PTEF13-B1	M	22	90	90	40	Private Pilot	0 hours	No
PTEF14-B4	M	19	91	91	15	Private Pilot	0-100 hours	No
PTEF15-B3	M	18	10	8	8	Student Pilot	1000+ hours	Yes - Flight Sim
PTEF16-B1	M	22	100	2	0	Private Pilot	100-500 hours	Yes - X Plane
PTEF17-B2	M	18	100	30	0	Private Pilot	0-100 hours	No
PTEF18-B3	M	21	125	15	1	Private Pilot	0-100 hours	Yes - Flight Sim
PTEF19-B4	M	22	135	10	0	Private Pilot	100-500 hours	Yes - Flight Sim

Part Task Fixed Feedback

			Flight Hours					
Participant Number	Gender	Age	Total	Last 12 Months	Last 90 Days	Ratings	Flight Sim Experience	737 Sim Experience
PTFF1-B1	M	20	130	60	90	Private Pilot Single and Multi	0-100 hours	No
PTFF3-B3	M	19	129	100	40	Private Pilot Single and Multi	0-100 hours	No
PTFF4-B1	M	19	100	50	20	Private Pilot	0-100 hours	No
PTFF5-B2	M	19	80	40	20	Private Pilot	100-500 hours	Yes - Flight Sim
PTFF6-B3	M	18	120	120	7	Private Pilot	0-100 hours	No
PTFF7-B4	M	18	115	0	0	Private Pilot	0-100 hours	No
PTFF8-B1	M	22	100	0	0	Private Pilot	0-100 hours	No
PTFF9-B2	M	19	60	20	5	Private Pilot	0-100 hours	Yes - Flight Sim
PTFF10-B3	M	23	80	80	15	Private Pilot	0-100 hours	Yes - Flight Sim
PTFF11-B3	M	37	61	0	0	Private Pilot	100-500 hours	Yes - Flight Sim
PTFF12-B4	F	28	140	40	20	Private Pilot	0 hours	No
PTFF13-B2	M	21	102	102	30	Private Pilot	0-100 hours	Yes - Flight Sim
PTFF14-B1	M	21	130	20	0	Private Pilot	0 hours	No
PTFF15-B4	M	19	40	40	40	Student Pilot	0-100 hours	No
PTFF16-B3	M	30	74	74	60	Private Pilot	0-100 hours	No
PTFF17-B2	M	20	120	20	0	Private Pilot	0-100 hours	No
PTFF18-B3	M	21	86	10	0	Private Pilot	0-100 hours	No

Part Task No Feedback

			Flight Hours					
Participant Number	Gender	Age	Total	Last 12 Months	Last 90 Days	Ratings	Flight Sim Experience	737 Sim Experience
PTNF1-B1	M	21	135	30	0	Private Pilot Single and Multi	100-500 hours	No
PTNF2-B2	M	18	101	50	25	Private Pilot	500-1000 hours	Yes - Flight Sim
PTNF3-B3	M	20	80	70	10	Private Pilot	0-100 hours	No
PTNF4-B1	F	21	94	30	15	Private Pilot	0-100 hours	No
PTNF6-B3	M	21	117	70	20	Private Pilot	0-100 hours	Yes - Flight Sim
PTNF9-B2	M	19	75	20	4	Private Pilot	1000+ hours	Yes - Flight Sim
PTNF10-B3	M	20	92	3.6	0	Private Pilot	100-500 hours	Yes - Flight Sim
PTNF11-B4	M	18	76	76	5	Private Pilot	100-500 hours	No
PTNF12-B3	M	22	55	55	0	Private Pilot	0-100 hours	Yes - Flight Sim
PTNF13-B1	M	21	120	75	15	Private Pilot	0-100 hours	No
PTNF14-B3	M	19	97	59	39	Private Pilot	0-100 hours	Yes - Flight Sim
PTNF16-B2	M	21	103	40	30	Private Pilot	0-100 hours	No
PTNF17-B1	M	25	120	10	2	Private Pilot	0-100 hours	Yes - Flight Sim
PTNF18-B2	M	19	66	65	15	Private Pilot	0-100 hours	No
PTNF19-B4	M	20	105	40	0	Private Pilot	500-1000 hours	No

Whole Task
Elaborative Feedback

			Flight Hours					
Participant Number	Gender	Age	Total	Last 12 Months	Last 90 Days	Ratings	Flight Sim Experience	737 Sim Experience
WTEF2-B2	M	21	100	90	20	Private Pilot	0-100 hours	No
WTEF3-B3	M	21	110	90	15	Private Pilot	0-100 hours	No
WTEF4-B1	M	18	70	70	5	Private Pilot	0-100 hours	No
WTEF5-B2	F	19	80	0	0	Private Pilot	0-100 hours	No
WTEF6-B3	M	18	45	20	15	Private Pilot	100-500 hours	No
WTEF7-B4	M	20	140	0	0	Private Pilot	0-100 hours	No
WTEF8-B1	M	18	75	35	5	Private Pilot	1000+ hours	Yes - Flight Sim
WTEF9-B2	M	21	87	27	20	Private Pilot	100-500 hours	No
WTEF10-B3	M	19	100	100	20	Private Pilot	0-100 hours	Yes - Flight Sim
WTEF11-B3	M	18	75	60	15	Private Pilot	100-500 hours	Yes - Flight Sim
WTEF12-B4	M	24	113	0	0	Private Pilot	0-100 hours	No
WTEF13-B1	M	21	75	20	10	Private Pilot	0-100 hours	No
WTEF14-B2	M	18	85	70	30	Private Pilot	0-100 hours	No
WTEF15-B3	M	21	105	60	10	Private Pilot	0-100 hours	No
WTEF16-B4	M	19	83	83	83	Student Pilot	0-100 hours	No
WTEF17-B1	M	19	73	73	30	Student Pilot	0-100 hours	Yes - Flight Sim
WTEF18-B4	M	19	110	30	6	Private Pilot	0-100 hours	No
WTEF19-B3	M	20	130	50	0	Private Pilot	0-100 hours	Yes - Flight Sim

Whole Task Fixed Feedback

Participant Number	Gender	Age	Flight Hours Total	Last 12 Months	Last 90 Days	Ratings	Flight Sim Experience	737 Sim Experience
WTFF2-B2	M	19	100	90	10	Private Pilot	0-100 hours	No
WTFF3-B3	M	19	70	70	30	Private Pilot	0-100 hours	Yes - Flight Sim
WTFF4-B1	M	19	95	75	25	Private Pilot	0-100 hours	No
WTFF6-B3	M	18	110	25	10	Private Pilot	0-100 hours	No
WTFF7-B4	M	20	140	45	10	Private Pilot	0-100 hours	No
WTFF8-B1	M	26	88	0	0	Private Pilot	0-100 hours	No
WTFF9-B2	M	19	80	80	0	Private Pilot	0-100 hours	No
WTFF10-B3	M	20	120	20	20	Private Pilot	500-1000 hours	Yes
WTFF11-B3	M	18	92	40	7	Private Pilot	0-100 hours	Yes - Flight Sim
WTFF12-B3	M	21	45	0	0	Student Pilot	0-100 hours	No
WTFF13-B1	M	21	115	20	8	Private Pilot	100-500 Hours	Yes - Flight Sim
WTFF14-B3	M	22	30	30	30	Student Pilot	0-100 hours	No
WTFF15-B3	M	21	120	70	20	Private Pilot	0 hours	No
WTFF16-B1	M	19	65	65	10	Student Pilot	100-500 hours	Yes - Flight Sim
WTFF18-B4	M	17	75	75	40	Private Pilot	0-100 hours	Yes - Flight Sim
WTFF19-B3	M	18	53	12	12	Private Pilot	100-500 Hours	No
WTFF20-B4	M	18	70	70	20	Private Pilot	0-100 hours	Yes - Flight Sim
WTFF21-B4	M	18	130	40	15	Private Pilot	0-100 hours	No

Whole Task No Feedback

			Flight Hours					
Participant Number	Gender	Age	Total	Last 12 Months	Last 90 Days	Ratings	Flight Sim Experience	737 Sim Experience
WTNF1-B1	M	18	130	70	0	Private Pilot	1000+ hours	Yes - Flight Sim
WTNF3-B3	M	18	115	40	20	Private Pilot	0-100 hours	No
WTNF5-B2	M	19	120	100	20	Private Pilot	100-500 hours	Yes - Flight Sim
WTNF6-B3	M	18	60	45	2	Private Pilot	500-1000	Yes - Flight Sim
WTNF8-B4	M	21	120	55	0	Private Pilot	100-500 hours	Yes - Flight Sim
WTNF10-B2	M	20	110	45	0	Private Pilot	0-100 hours	No
WTNF11_B3	F	21	100	0	0	Private Pilot	0-100 hours	No
WTNF12_B4	M	19	70	70	25	Private Pilot	0-100 hours	No
WTNF13-B3	M	24	85	85	25	Private Pilot	0-100 hours	No
WTNF14-B4	M	22	110	50	10	Private Pilot	0-100 hours	Yes - Flight Sim
WTNF15-B2	F	21	140	50	0	Private Pilot	0-100 hours	No
WTNF16-B3	M	20	120	90	15	Private Pilot	0-100 hours	Yes - Flight Sim
WTNF17-B4	M	22	100	30	10	Private Pilot	100-500 hours	No
WTNF19-B3	M	20	70	70	30	Private Pilot	500-1000 hours	Yes - Flight Sim
WTNF20-B4	M	21	120	10	10	Private Pilot	0-100 hours	Yes - Flight Sim
WTNF21-B1	M	22	91	34	1	Private Pilot	0-100 hours	Yes - Flight Sim

Final Test Results

Part-Task Elaborative Feedback

Participant Number	Landing Altitude	Altitude	Bank Angle	Brakes	Course	Flaps	Gear	Glide slope	Heading	Speed	Cumulative
PTEF1-B1	442	10	0	0	20	0	0	50	0	100	180
PTEF2-B2	442	10	10	10	0	0	50	50	0	200	330
PTEF4-B1	442	30	0	0	50	0	0	30	0	210	320
PTEF5-B2	442	0	0	0	0	0	0	10	0	30	40
PTEF6-B3	442	0	0	0	0	0	0	10	0	90	100
PTEF7-B3	442	60	0	0	0	0	0	90	0	140	290
PTEF8-B2	442	90	0	10	40	20	0	10	0	190	360
PTEF10-B2	455	60	0	0	0	0	0	0	0	30	90
PTEF11-B1	434	70	0	0	40	0	0	70	0	200	380
PTEF12-B4	434	70	0	0	40	0	0	70	0	200	380
PTEF13-B1	1070	20	10	0	0	0	0	0	20	30	80
PTEF14-B4	442	30	0	0	60	0	10	80	0	180	360
PTEF15-B3	442	10	0	0	0	0	10	0	0	80	100
PTEF16-B1	442	10	0	10	0	0	0	10	0	170	200
PTEF18-B3	442	10	0	10	0	0	0	30	0	110	160
PTEF19-B4	442	10	10	10	0	0	10	10	0	20	70

Part-Task Fixed Feedback

Participant Number	Landing Altitude	Altitude	Bank Angle	Brakes	Course	Flaps	Gear	Glide slope	Heading	Speed	Cumulative
PTFF1-B1	442	70	10	0	0	0	0	10	0	260	350
PTFF3-B3	442	10	0	10	90	100	0	20	0	180	410
PTFF4-B1	442	0	10	0	30	0	0	30	0	280	350
PTFF5-B2	442	0	0	0	0	0	0	10	0	10	20
PTFF6-B3	442	40	0	0	0	0	0	10	0	50	100
PTFF8-B1	442	0	0	0	0	0	0	10	0	50	60
PTFF9-B2	442	0	0	0	0	0	0	0	0	80	80
PTFF10-B3	442	10	0	0	0	0	0	10	0	20	40
PTFF11-B3	442	4	0	0	0	0	10	10	0	50	110
PTFF13-B2	442	0	0	0	0	0	0	10	0	80	90
PTFF14-B1	442	60	10	0	40	0	0	40	40	140	330
PTFF15-B4	442	110	0	0	0	0	10	0	0	180	300
PTFF16-B3	291	20	0	0	20	0	0	30	0	270	340
PTFF17-B2	442	0	20	0	40	0	0	10	0	60	130
PTFF18-B3	442	60	0	0	0	0	0	0	0	60	120

Part-Task No Feedback

Participant Number	Landing Altitude	Altitude	Bank Angle	Brakes	Course	Flaps	Gear	Glide slope	Heading	Speed	Cumulative
PTNF1-B1	442	0	0	0	80	0	0	10	0	30	120
PTNF2-B2	442	0	0	0	80	0	0	10	0	30	120
PTNF3-B3	442	30	0	10	50	0	0	10	0	170	270
PTNF4-B1	4183	60	0	0	0	0	0	0	0	130	190
PTNF6-B3	442	20	0	0	0	0	0	10	0	80	110
PTNF9-B2	442	0	0	0	0	0	0	40	0	120	160
PTNF11-B4	433	0	10	0	0	10	0	10	0	200	230
PTNF12-B3	442	0	0	0	30	0	0	10	0	40	80
PTNF13-B1	564	10	10	0	0	0	0	0	0	140	160
PTNF14-B3	442	0	30	0	20	0	10	10	0	90	160
PTNF18-B2	442	10	10	0	0	0	0	20	0	70	110

Whole-Task Elaborative Feedback

Participant Number	Landing Altitude	Altitude	Bank Angle	Brakes	Course	Flaps	Gear	Glide slope	Heading	Speed	Cumulative
WTEF2-B2	442	0	0	0	0	0	0	10	0	40	50
WTEF3-B3	432	30	10	0	0	10	0	30	0	120	200
WTEF6-B3	442	30	10	0	0	0	0	10	0	130	180
WTEF7-B4	442	50	0	10	0	0	0	20	0	120	200
WTEF8-B1	442	40	10	0	0	0	0	10	0	60	120
WTEF9-B2	442	20	0	0	0	0	0	10	0	90	120
WTEF10-B3	442	10	0	10	0	0	0	10	0	40	70
WTEF11-B3	442	10	10	0	0	0	0	10	0	80	110
WTEF12-B4	442	20	10	20	0	0	0	20	0	80	150
WTEF13-B1	442	20	0	0	0	0	0	10	0	30	60
WTEF17	442	0	10	0	0	0	40	10	0	210	270
WTEF18	442	0	0	0	0	0	120	10	0	50	180

Whole-Task Fixed Feedback

Participant Number	Landing Altitude	Altitude	Bank Angle	Brakes	Course	Flaps	Gear	Glide slope	Heading	Speed	Cumulative
WTFF2-B2	442	0	0	10	20	0	0	10	0	50	90
WTFF3-B3	442	20	0	0	40	0	0	0	0	90	150
WTFF4-B1	442	10	0	0	0	10	310	10	0	120	460
WTFF7-B4	442	20	0	0	0	0	0	20	0	90	130
WTFF8-B1	442	0	10	0	0	0	0	20	0	50	80
WTFF9-B2	442	10	10	0	10	0	0	10	0	190	230
WTFF10-B3	442	0	0	0	0	0	0	10	0	10	20
WTFF11-B3	442	0	0	0	10	0	0	10	0	130	150
WTFF12-B3	442	0	0	10	0	0	0	30	0	210	250
WTFF13-B1	442	20	0	10	0	0	0	20	0	10	60
WTFF14-B3	442	0	0	0	0	0	0	30	0	30	60
WTFF15-B3	442	0	0	0	0	0	0	60	0	60	120
WTFF16-B1	442	0	0	0	0	0	0	0	0	20	20
WTFF18-B4	442	0	0	0	0	0	0	10	0	50	60
WTFF19-B3	442	0	0	0	30	0	0	10	0	60	100
WTFF20-B4	442	0	0	0	0	0	0	20	0	30	50

Whole-Task No Feedback

Participant Number	Landing Altitude	Altitude	Bank Angle	Brakes	Course	Flaps	Gear	Glide slope	Heading	Speed	Cumulative
WTNF1-B1	442	0	0	0	0	30	0	10	0	60	100
WTNF8-B4	442	0	0	10	0	0	0	10	0	90	110
WTNF12_B4	442	0	0	0	20	0	0	70	0	150	240
WTNF16-B3	442	20	0	0	0	10	0	20	0	190	240
WTNF17-B4	440	0	0	0	0	0	0	0	0	60	60
WTNF19-B3	442	40	10	10	0	0	0	10	0	120	190
WTNF20-B4	442	50	0	0	0	0	0	20	0	120	190