Southeastern University
FireScholars

College of Education

Spring 2018

A STUDY OF FLIGHT SIMULATION TRAINING TIME, AIRCRAFT TRAINING TIME, AND PILOT COMPETENCE AS MEASURED BY THE NAVAL STANDARD SCORE

Aaron D. Judy
Southeastern University - Lakeland

A STUDY OF FLIGHT SIMULATION TRAINING TIME, AIRCRAFT TRAINING TIME, AND PILOT COMPETENCE AS MEASURED BY THE NAVAL STANDARD SCORE

By

AARON JUDY

A doctoral dissertation submitted to the
College of Education
in partial fulfillment of the requirements
for the degree Doctor of Education
in Organizational Leadership

Southeastern University
April 9, 2018

A STUDY OF FLIGHT SIMULATION TRAINING TIME, AIRCRAFT TRAINING TIME, AND PILOT COMPETENCE AS MEASURED BY THE NAVAL STANDARD SCORE

By

AARON JUDY

Dissertation Approved: Ap 9, 2018

Patty LeBlanc, Ph.D., Dissertation Chair

Tom Gollery, Ed.D., Committee Member

Ryan Wallace, Ed.D., Committee Member

Jim Anderson, Ph.D., Dean, College of Education

DEDICATION

I would like to thank all those that have served in our military; your sacrifice and dedication to our country helped make this research study possible. I would like to thank my advisors and classmates in the doctoral program. Your support, friendship, and encouragement throughout this journey has been incredible. I am forever grateful to both my grandfathers who served our great country in WWII and passed the love of aviation down to me. Thank you for all the love and encouragement in my life, career, and academic journey.

I am also so grateful for my dad, Anthony Judy, and mom, Charlotte Judy, who taught me from an early age that by keeping God first and never giving up, I would accomplish my goals. Mom, I have come a long way from when I first got off the school bus in the 1^{st} grade and said, "I'll be glad when I don't have to go to school again."

To my wife, Christin Judy, who is the real educator as an elementary teacher, thank you for your never ceasing support and encouragement. You make every day exciting, and I appreciate all your patience and love while I continuously worked for hours in the evenings and on the weekends. You always said "You can do it" when I was tired and did everything you could to ensure I was successful. You are the love of my life and my best friend.

Finally, I thank God and give Him all the glory, for nothing would be possible without His strength and wisdom.

ACKNOWLEDGMENTS

I want to thank my United States Navy family at the Naval Air Warfare Centers and United States Special Operations Command. Your professional support, advice, and belief in me to carry the responsibility to help provide the best aircraft systems and training systems to our airmen and sailors is a privilege. Thank you to CDR Andrew McLean at Training Wing TWO in Kingsville, Texas, for letting me shadow you during a week of Navy T-45 flight simulation and flight training. The insight you provided into the T-45 flight training program was instrumental. I appreciate all the help from the Chief of Naval Air Training program. Thank you to Dr. Joe Geeseman for helping me access data and providing valuable feedback during the research and dissertation process.

Thank you, Dr. Cindy Campbell, for serving as my editor and providing valuable edits to help make my dissertation complete. You spent many hours working on my dissertation and were always there to answer any questions.

Thank you, Dr. Ryan Wallace, for serving as my committee member and sharing all your advice during this program. As a professor at Polk State College and now at Embry-Riddle Aeronautical University, you encouraged me to keep going and provided valuable information to help get me to the finish line.

Dr. Gollery, my research methodologist, your humbleness and dedication to the doctoral program is inspiring. You were always there day or night to answer my questions about data analysis. Thank you for all your time and investment in helping me succeed. You are a blessing to Southeastern University and brilliant!

Dr. LeBlanc, my chair, thank you so much for taking me on as your dissertation student. Words are hard to describe how much I appreciate you and your dedication to students, teaching,

and learning. I have learned so much about education, research, and writing under you, and I will always be grateful. You are a real leader in education!

ABSTRACT

The purpose of the study was to investigate the relationships between US Navy T-45C flight simulation training time, actual aircraft training time, and intermediate and advanced jet pilot competence as measured by the Naval Standard Score (NSS). Examining the relationships between US Navy T-45C flight simulation time and actual aircraft flight time may provide further information on flight simulation training versus actual aircraft training to aviation authorities, flight instructors, the military aviation community, the commercial aviation community, and academia. The study was non-experimental, correlational, causal-comparative with an emphasis upon the establishment of mathematic and predictive relationships using archival data from the Chief of Naval Air Training (CNATRA) Training Information System (TIMS) database. CNATRA aircraft hours, flight simulation hours, and NSS scores of intermediate and advanced flight students from 2015 to 2017 were analyzed and compared. Actual aircraft time was found to be a significant predictor of NSS scores for both intermediate and advanced pilot trainees. Implications of the study include recommendations for future research and strategies to improve flight simulation in pilot training.

Keywords: flight training, flight instruction, pilot training, aircraft training, military training, simulation, flight simulation, simulator training, US Navy, learning, memory, cognition

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Military pilot training requires an enormous investment of time, energy, and resources. In fiscal year 2017, the United States (US) Department of Defense spent approximately 1.6 billion dollars on flight training (Department of Defense, 2016). The Department of the Navy is responsible for training both US Navy and US Marine pilots who are trained at various Training Air Wings located across the US. Military pilot training is extensive, costly, and in high demand. Admiral Moran, the Vice Chief of Naval Operations who is responsible for pilot training readiness, testified before the US House Armed Services Committee that "there are three main drivers of our readiness problems: persistent, high operational demand for naval forces; funding reductions; and consistent uncertainty about when those reduced budgets will be approved" (US Congress, 2017, p. 2).

With the advent of computer-based technologies, pilot training via computer simulations became standard practice in military pilot training and has been, on the whole, very effective and efficient in developing the skills of naval pilots (Department of the Navy, 2010). Both the United States Navy (USN) and the United States Marines Corps (USMC) are very interested in research designed to determine the optimal amount of training time necessary to achieve pilot expertise in various aircraft. A major question posed by the Navy and the USMC focuses on the feasibility of reducing actual aircraft training time, which is labor and cost intensive, through the use of flight simulation training. In 2010, the Navy estimated that 61% of flight training should be actual aircraft training (Department of the Navy, 2010). The Navy then asked whether enhanced simulation capabilities could reduce actual aircraft training time to 44% without reducing training and pilot effectiveness (Department of the Navy, 2010). This important question was the impetus for the current study. Comparing actual aircraft training time and flight simulation training time to overall pilot competence would make an important contribution to policy and decision-making among the nation's military forces.

The USN and USMC currently train fixed-wing pilots through Command Training Wings located across the US. Primary training, which is the introduction to jet training, is conducted in a Beechcraft T-6B Texan aircraft located at Training Wing FOUR in Corpus Christi, Texas, or at Training Wing FIVE located in Milton, Florida. Primary training takes approximately 28 weeks to complete. Upon completion of primary flight training, jet pilots move into intermediate training, followed by advanced training before earning their Wings of Gold. Intermediate and advanced flight training is conducted in a Boeing T-45C Goshawk aircraft at Training Wing ONE in Meridian, Mississippi, or at Training Wing TWO in Kingsville, Texas. Intermediate training takes approximately 27 weeks to complete, and advanced training takes approximately 25 weeks to complete. Pilots who graduate from Training Wing ONE and TWO supply the operational needs of the Navy and USMC for F/A-18 aircraft and next-generation F-35 jet aircraft pilots.

Military flight training currently consists of a combination of classroom instruction, simulator time, and aircraft flight time. Existing USN and USMC student pilot training data are available from the Chief of Naval Air Training (CNATRA). The dataset contains intermediate and advanced student pilot-level data on simulator training time, aircraft training time, student scores on simulator and aircraft training events, and overall pilot effectiveness as measured by the Naval Standard Score (NSS). The NSS score refers to overall pilot proficiency as determined by the flight instructor.

The current study was designed to investigate intermediate and advanced T-45C Goshawk aircraft training hours and T-45 Operational Flight Trainer (OFT) simulation training hours and their relationships to the Naval Standard Score of intermediate and advanced student pilots at Training Wing ONE and TWO. The researcher compared flight simulation hours, actual aircraft training hours, and their relationships to scores on the NSS of USN and USMC student aviators.

Background of the Study

The US Navy is known for adopting innovative and integrated training solutions to increase operational readiness. The Chief of Naval Operations (CNO) established a simulator training strategy in 2014 that stated, "To the maximum extent practicable, live training should be completed by simulators or fleet synthetic training where training effectiveness, safety, and operational readiness are not compromised" (Department of the Navy, 2014, p. 2). Therefore, mapping essential aircraft training tasks to simulator devices is an important step to ensuring that operational readiness is not compromised. However, research suggests that flight simulation, though suitable for training some of the myriad tasks required, may not be the best method for training certain critical aircraft tasks, such as tasks requiring significant mental problem-solving and high levels of performance (Beaubien, Stacy, Wiggins, & Lucia, 2016).

Pilot Learning and Performance

Beaubien et al. (2016) stated that learning and performance can be identified and measured in multiple ways. Learning can be defined as a change in knowledge, skills, or understanding across environments and time. Performance can be defined as the quality, rate, or accuracy of a response at a specific point in time (Beaubien et al., 2016). Student pilots often practice a task to the point of automaticity and eventually reach a plateau. However, the extent of pilot learning is often not easily demonstrated after a period of disuse, which can occur after pilots graduate from pilot school and discontinue the rigorous training schedule. Experimental studies of student pilots suggest that under controlled conditions, pilots who practice higher levels of overlearning tend to show less decay and faster reacquisition of learning (Beaubien et al., 2016).

Beaubien et al. (2016) conducted a study of Navy F/A-18 pilots that investigated methods to measure the ways learning takes place during simulated carrier landings. The authors studied 15 Navy F/A-18 pilots who flew 24 landing passes in a high-fidelity simulator. Measures of Performance (MOPs) were analyzed for each landing and assessed during the last 18 to 23 seconds of the final approach to landing. The MOP scores were then averaged, and each pilot was assigned a score. Based on the simulator analysis, the MOP scores demonstrated that flight performance improved over time. However, the Measures of Learning scores (MOLs) revealed that auditory, visual hints, and cues provided by the flight simulator actually resulted in reduced mean scores on final approach to landing in actual aircraft. In other words, negative transfer of learning occurred from the simulator to actual pilot performance in the aircraft. Based on the results of this study, both MOP and MOL scores should be assessed when conducting simulated flight training or when improving flight simulations (Beaubien et al., 2016).

Problem-solving, learning, memory, and cognition are all key components of flight training, and a plethora of research studies have been conducted in a wide variety of disciplines. Classifying the cognitive demands of pilots using a simulated device is an important method of describing and measuring learning during flight training to help ensure proficiency (Hoke, Reuter, Romeas, Montariol, Schnell, & Faubert, 2017). Hoke et al. (2017) conducted a study using a Cognitive Assessment Tool Set (CATS) system, which was worn under a pilot's flight suit. The CATS system did not interfere with flying tasks and provided a real-time assessment of the cognitive workload during flight training. The study's sample included 10 low-time (100 to 300 flight hours) pilots between the ages 20 and 25 who held a valid US private pilot certificate with a Class III medical certification. A Class III medical certification is granted by a designated Federal Aviation Administration (FAA) medical examiner for recreational and private pilots. The study used Neurotracker (NT) to evaluate the pilot's cognitive workload when piloting an L-29 jet trainer aircraft and an L-29 simulator. NT is a scientific instrument developed to help improve perceptual-cognitive abilities among athletes. NT, also referred to as Three-Dimensional Multiple Object Tracking (3D-MOT), isolates a number of mental skills used for reading and training and closely monitors the brain's processes during complex motion. Faubert and Sidebottom (2012) developed the 3D-MOT methodology to stimulate and measure brain networks that work together during motion processing, attention processing, and working memory. Using NT methodology, the Hoke et al. (2017) study revealed that flying, whether in an aircraft or a flight simulator, is a cognitively demanding task. However, the results of the study showed that the pilots' maneuvers as measured by NT were more cognitively demanding when flying the L-29 jet aircraft than when flying in a L-29 simulator (Hoke et al., 2017). The current study provides important evidence to address the questions that the USN and USMC seek to answer regarding flight simulation versus actual aircraft time and overall pilot effectiveness.

A study of simulator training versus aircraft pilot training was funded and conducted by the FAA to investigate the transfer of learning after upset-recovery flight training in a low-cost flight simulation device (Rogers, Boquet, Howell, & DeJohn, 2007). A flight upset occurs when an aircraft enters into loss-of-control in-flight and impacts the ground. Upset-recovery is extremely important in aviation, and the inability to recover from upsets has been responsible for multiple fatal air accidents (Rogers et al., 2007). Pilots normally conduct upset-recovery training in flight simulation devices, but little research has been conducted to determine whether simulation training transfers to the ability of a pilot to regain loss of control in an actual airplane (Rogers et al., 2007). The FAA study included 60 pilot trainees at Embry-Riddle Aeronautical University (ERAU); 30 pilots were randomly assigned to the experimental group, and 30 were assigned to a control group. The 30 pilots in the experimental group participated in classroom academic instruction on upset-recovery using Microsoft Flight Simulator 2002 on a desktop computer. The 30 pilots in the control group did not participate in classroom instruction using Microsoft Flight Simulator, 2002. The experimental and control group pilots later participated in the actual flight test in an E33C Beech Bonanza aircraft that was fitted with a video recording device and a flight data recording (FDR) device. Many technical difficulties with the instrumentation occurred during the experiment in the E33C Beech Bonanza, but some of the experiment's results were worth noting. Assessment of the video and audio data of both the experimental and control groups revealed that there were no significant differences between the two groups' upset-recovery performance during actual flight. The authors of the study stated, "It may well be the case that simulator-trained pilots need attitude [positional information] flight experience in an actual airplane to hone their simulator-based upset-recovery skills to an acceptably high level" (Rogers et al., 2007, p. 16). This research study suggests that effective flight training can be conducted using a combination of both aircraft training and flight simulation training. However, the question remains: How much simulator training can actually replace aircraft training without negatively impacting pilot effectiveness?

Problem and Purpose Statements

The US Navy desires to increase flight simulation time and decrease actual aircraft training time in order to reduce costs and to increase the lifespan of aging aircraft. However, research suggests that flight simulation training, though highly efficient, lacks the fidelity needed to emulate many critical flying tasks (Beaubien et al., 2016). The purpose of the current study is to investigate the relationships between flight simulation training, actual aircraft training, and intermediate and advanced jet pilot competence as measured by the Naval Standard Score.

Overview of Methodology

The research design of the study was non-experimental, correlational, causal-comparative, and predictive research using archival data**.** The sample selection was purposive and provided by CNATRA's Training Information Management System (TIMS) database. The dataset was purposive; the data were selected for only intermediate and advanced flight students who trained on the T-45C Goshawk aircraft and T-45C Operational Flight Trainer (OFT). Participants included USN and USMC intermediate and advanced T-45C flight students from years 2015 to 2017. The sample size consisted of 358 intermediate flight students and 334 advanced flight students from Training Wings ONE and TWO. CNATRA aircraft hours and flight simulation hours were compared and used to predict NSS scores. Although flight students are required to complete a certain number of actual aircraft training hours and flight simulation hours, students can request more unproctored time in the simulator. This study was designed to examine actual aircraft hours and flight simulation hours and their relationships to the NSS of intermediate and advanced student pilots.

Research Questions

In order to address the stated research problem, the following research questions and hypotheses were posed:

Q1. Which is the best predictor of the NSS scores of intermediate pilot trainees: flight simulation time or actual aircraft time?

Q2. Which is the best predictor of the NSS scores of advanced pilot trainees: flight simulation time or actual aircraft time?

Q3. Are there any significant differences between intermediate and advanced pilot trainees' flight simulation time and actual aircraft time?

Hypotheses

$H_0$1: Flight simulation training time and actual aircraft training time are not significant predictors of NSS scores of intermediate pilot trainees.

H_o2: Flight simulation training time and actual aircraft training time are not significant predictors of NSS scores of advanced pilot trainees.

H_o 3: There is no significant difference between mean flight simulation time and mean actual aircraft time of intermediate and advanced pilot trainees.

To address the first and second research hypotheses, multiple linear regression was used to determine whether there were significant predictor(s) of the NSS scores. The independent variables included flight simulation time and actual aircraft time; the dependent variable was the NSS scores of intermediate pilot trainees and advanced pilot trainees. Using the Fisher *r* to *z* transformation test statistic*,* correlations between flight simulation time and actual aircraft time on NSS were evaluated. Cohen's *q* test statistic was used to evaluate the comparative relational effect between flight simulation time and actual aircraft time. The statistical significance of the first and second hypotheses utilized the .05 alpha level as the threshold for statistical significance.

Due to the non-normal distribution of the data arrays, the Mann-Whitney U test was used to evaluate the statistical significance of differences in mean ranks for training type (flight simulation time and actual aircraft time) and by training phase (intermediate and advanced) to address Research Question three. Cohen's *q* statistic was used to evaluate the magnitude of differences of mean rankings (effect size) in the respective comparisons. The alpha level of .05 was utilized as the threshold value of statistical significance.

Limitations of the Study

Although this study was intended to provide information on the relationships of flight simulation time, actual aircraft time, and pilot performance, there were limitations to the study. The researcher assumed that either flight simulation training time or actual aircraft training time were predictive of NSS scores, but that the relationships were unknown. The sample was purposive and drawn from two branches of military service, the USN and USMC, and may not be representative of all military or civilian agencies. The dataset was derived from the T-45C Goshawk aircraft and flight simulator, the T-45C OFT. CNATRA provided the archival dataset for years 2015 through 2017. The years 2015-2017 were utilized to control for changes in aircraft and simulation technologies. At the time of this study, the T-45C pilot training program experienced problems with the On-Board Oxygen Generation Systems (OBOGS); the analysis of the pilot data in this study does not take into account possible changes in the morale of the CNATRA T-45C pilots or instructors in the flight training program as a result of these problems.

Definitions of Key Terms

Advanced Training

Advanced training is the CNATRA phase of training conducted in a T-45C Goshawk aircraft and includes advanced training in tactics and aircraft carrier landing qualifications (Naval Air Training Command, 2014).

Aircraft Training

Aircraft training is training flown in a real aircraft. Sometimes aircraft training can be referred to as "live" or "real training".

Augmented Reality

A visual technology that superimposes a computer-generated image on a user's view in the real world or in a simulation.

Automaticity

Automaticity refers to a task which occurs automatically after considerable practice and with little cognitive thought.

eLearning

eLearning refers to digital learning normally provided online or delivered by a digital method such as a computer. eLearning is also sometimes called distance learning.

Fleet Synthetic Training

Fleet Synthetic Training (FST) events are computer-assisted exercises provided by training agents during various phases of pilot training that bring together multiple interoperable training devices (Department of the Navy, 2014).

Flight Simulator

Flight simulators are computer-based training consoles that include a cockpit and/or emergency procedure trainer (Naval Air Training Command, 2014). Sometimes a flight simulator is referred to as an Operational Flight Trainer (OFT).

Flight Simulation Training

Flight simulation training is digital training that is conducted in a computer-based simulator to emulate actual aircraft flight tasks. Sometimes flight simulation training may be referred to as "virtual training".

Intermediate Training

Intermediate training is the CNATRA phase of training conducted in a T-45C Goshawk aircraft for flight students on track to becoming jet strike pilots (Naval Air Training Command, 2014).

Learning

Pilot Learning is defined as a change in knowledge, skills, or understanding across environments and time (Beaubien et al., 2016).

Naval Standard Score (NSS)

The NSS is used by Navy personnel to calculate and to correct for potential non-normality in the distribution of the Phase Aggregate Score (PAS). PAS is a comparative ranking of flight students' performance on a group of training events compared to rankings of previous flight students' performance for the same set of training events (Naval Air Training Command, 2014). More information on the NSS is provided in Chapter IV. The equation used to calculate NSS scores is presented below.

$$NSS = 50 + 10 * (PAS - MPAS/SDPAS)$$

Part Task Trainer

A part task simulator trains only a limited number of aircraft partial tasks.

Performance

Pilot performance is defined as the quality, rate, or accuracy of a response at a specific point in time (Beaubien et al., 2016).

Proficiency

Pilot proficiency is defined as the level of performance indicated by successful accomplishment of a task as described in a curriculum, flight training instruction, or a manual (Naval Air Training Command, 2014).

Primary Training

Primary training is the CNATRA phase of training conducted in a T-6B aircraft and is the foundational phase of flight training (Naval Air Training Command, 2014).

Reacquisition

The action or process of acquiring a learned task after its loss.

Simulation Fidelity

Flight Simulation Fidelity is defined as the degree of similarity between a simulated task and an operational task that is simulated (Hays, 2006). The degree of fidelity can be explained as the interaction of two characteristics: the physical characteristics of the flight simulation (e.g., visual, spatial, and auditory) and the functional characteristics of the flight simulation (e.g., information and responses provided by the simulation) (Hays, 2006).

Training Effectiveness

Training effectiveness has a plethora of meanings across various organizations and environments and is important to define for this specific study. Kirkpatrick's (1976) seminal study defines training effectiveness as outcomes of reactions, learning, behavior, and organizational results; his definition is utilized in this study.

Training Event

Training events are scheduled periods and tasks of flight instruction. A flight training event can be held in an academic or laboratory classroom, a flight simulator, or an actual aircraft (Naval Air Training Command, 2014).

Virtual Reality

A computer-generated simulation or environment that provides physical real-time interaction using special electronic equipment (e.g., sensors, gloves, and helmets).

II. REVIEW OF LITERATURE

Pilot training research has been ongoing for the last century, and the results have had a great influence on how pilots train and learn today. Training research encompasses a larger body of literature and helped define the parameters of this study although the study focused on military flight training.

This chapter serves as a review of relevant literature on cognitive theory, learning, memory, training research, training effectiveness, and studies related to commercial and military flight training. Learning by adults is emphasized in this literature review although someone at the age of 16 can obtain a private pilot's license (FAA, 2018); however, this study was primarily concerned with flight students aged 18 or older. Career flight training typically occurs at the age of 18 either through a collegiate flight school or the military.

Flight training programs normally use a combination of actual aircraft flight training, flight simulation training, one-on-one ground instruction, and classroom instruction. Jacobs, Prince, Hays, and Salas (1990) conducted a meta-analysis of the effectiveness of flight simulations and concluded that flight simulation training produces superior training relative to aircraft-only training. However, the optimal combination of all three instructional methodologies (aircraft, simulation, and classroom) and their impacts on pilot cognition, working memory, long-term memory, and retention are the subjects of important research needs in the military.

The Navy uses training and readiness (T&R) matrices to determine the tasks that should be conducted in flight training and their priorities (Schank, Thie, Graff, Beel, & Sollinger, 2002). The T&R matrices are used to identify a task number that specifies whether the task can be flown, conducted in a simulator, or both (Schank et al., 2002). Typically, a thorough job task and training methodology analysis is conducted by instructional design psychologists to determine whether tasks should be conducted in classroom training, flight simulation training, or actual aircraft training (Foster, Melon, & Phillips IV, 2007). The training methodology analysis includes a job analysis and task list, learning objectives with conditions and standards, and a media analysis of simulation, aircraft, or classroom instruction (Foster et al., 2007). The media analysis consists of four steps: (a) identify learning domains by objective, (b) identify and determine stimulus and response requirements, (c) identify the degree of physical and functional fidelity required, and (d) match the data to the best delivery method (simulated versus actual aircraft), taking into consideration the learning objectives and the relative cost (Foster et al., 2007). Based on these analyses, the T&R matrix can specifically identify training tasks and recommended training flight hours, simulator hours, and classroom instruction. These task analyses are critically important to the overall decision-making of flight instructors and the military.

In 2010, the Navy estimated that 61% of flight training should be conducted in an actual aircraft (Department of the Navy, 2010). The Navy then asked whether it would be possible to enhance flight simulation capabilities so that a minimum of 44% of pilot training could be flown in actual aircraft (making flight simulation time 56%) without reducing pilot effectiveness (Department of the Navy, 2010). This study sought to better understand the relationships between pilot cognition, learning, memory, cognitive behavior, training, and pilot effectiveness in order to advise the Navy regarding optimal ratios of flight simulation training to actual aircraft training.

Theoretical Framework

Bloom's Taxonomy of the Cognitive Domain is one of the most well-known models for describing thinking, learning, and instruction (Palmer, 2001). Bloom (1956) conducted extensive research leading to the identification of six levels of human thought from lowest to highest complexity: *Knowledge*, *Comprehension*, *Application*, *Analysis*, *Synthesis*, and *Evaluation* (Bloom, 1956). Each of the levels of thinking includes and builds on the previous levels in the hierarchy. Over time, the taxonomy of the cognitive domain came to be known as Bloom's Taxonomy. A visual depicting the six levels of Bloom's Taxonomy and their relationships are shown in Figure 1.

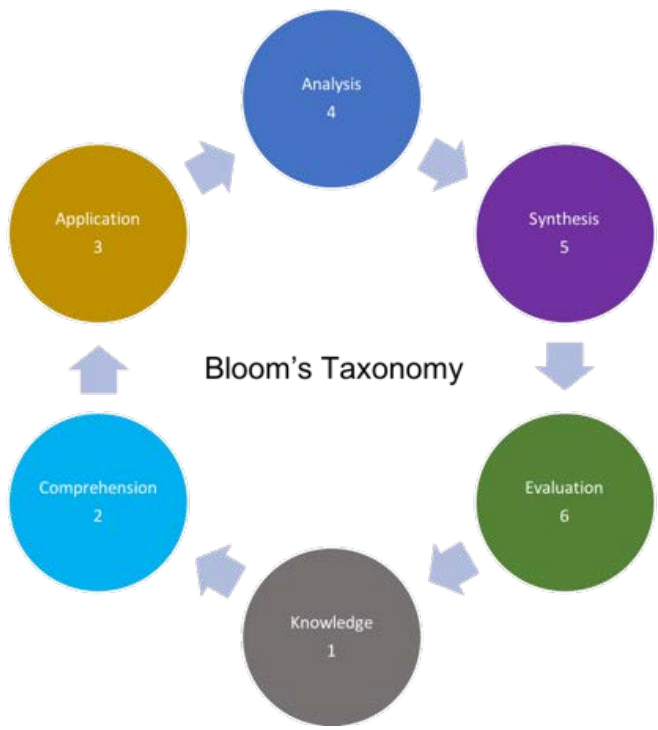

Figure 1: Bloom's Taxonomy (Adapted from FAA, 2008, p. 2-5).

The lowest level of thinking described in the hierarchy is *Knowledge.* At this level, learners are challenged to store and recall information and principles in the form in which they were originally learned. Bloom's second level is *Comprehension*, in which learners comprehend and interpret information based on prior knowledge gained in the first level and on the connections made from prior knowledge. These first two levels of thinking are considered to be lower-order levels of thinking and reasoning and typically occur when the learner is exposed to new information. In the case of adult learning, this new information may be quite complex, as in the case of a pilot's first exposure to emergency procedure recoveries.

Bloom's third level of thinking deals with *Application* of knowledge and skills*,* in which learners apply information in similar or non-similar situations to those in which learning first occurred. This level of thinking requires the ability to reason abstractly. The application level of the hierarchy is considered to be the first level of higher-order thinking. The fourth level in the hierarchy is *Analysis,* which may require learners to compare and classify information, make predictions, form a hypothesis, and structure a problem statement. Bloom's fifth level describes *Synthesis,* in which learners combine knowledge, thoughts, or ideas to create a product, idea, or plan that is new. Finally, Bloom's sixth level is *Evaluation,* in which learners assess or critique information based on all that they know, comprehend, and can apply, analyze, and synthesize regarding a specific task, problem, or idea.

The robust and highly descriptive nature of Bloom's Taxonomy was readily adopted by educators of all age groups and disciplines as a way to focus instruction on higher order thinking skills and integration of prior knowledge. As a result, the process of teaching and the development of instructional curricula were radically impacted by the pedagogical implications of the taxonomy. Some of the cognitive tasks thinkers may use at each of the Bloom levels is depicted in Figure 2.

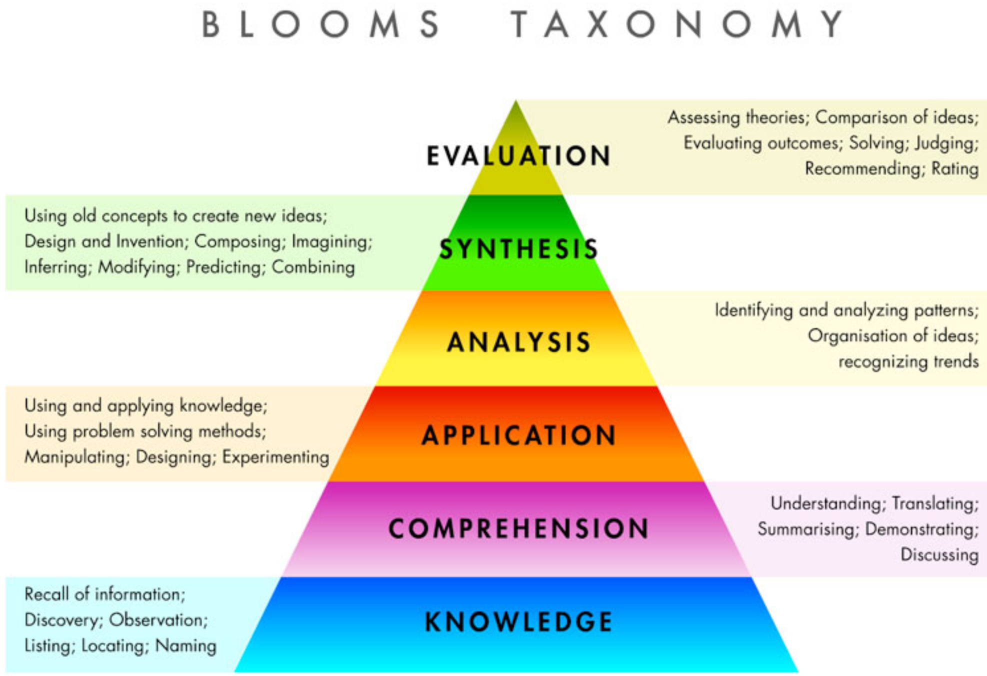

Figure 2: Samples of Cognitive Tasks in Bloom's Hierarchy

Bloom's Taxonomy provided educators with a model of instruction that could be prescribed, described, and measured. In fact, Bloom's work is one of the best-known examples of research evidence that informed educational practice (Palmer, 2001). The taxonomy remains as a comprehensive model for describing both lower-order and higher-order thinking, and it provides an excellent framework for understanding the higher-order thinking required by pilots when flying an aircraft.

Bloom's Revised Taxonomy

Educational researchers have revised Bloom's original taxonomy to focus on the educational implications of cognitive thinking (Sousa, 2017). Anderson and Krathwohl (2001) revised the original taxonomy to reflect six levels of human thought: *Remember*, *Understand*, *Apply*, *Analyze*, *Evaluate*, and *Create*. A comparison of Bloom's original taxonomy and the revised taxonomy is depicted in Figure 3.

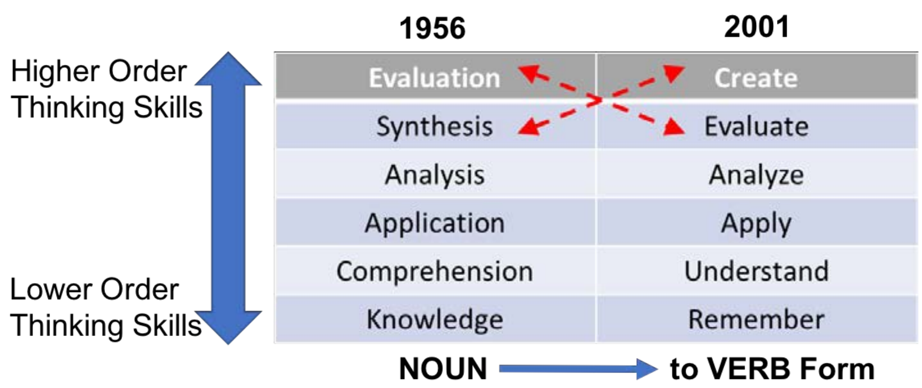

Figure 3: Bloom's Revised Taxonomy

Sousa (2017) describes the rationale for the changes to the original model. *Remember* replaced *Knowledge* to more accurately describe the brain's recall processes. *Comprehension* was re-named *Understand* since instructors use the word "understand" more commonly. *Application*, *Analysis*, and *Evaluation* were changed to *Apply*, *Analyze*, and *Evaluate*. Finally, *Synthesis* changed places with *Evaluation*, which became *Create*. The addition of *Create* was based on recent studies in cognitive neuroscience suggesting that planning and producing requires complex thinking skills (Sousa, 2017). However, pilots typically do not *Create* in the cockpit; therefore, Bloom's original taxonomy of the cognitive domain was chosen to describe the conceptual framework for this study. In addition, the original hierarchy developed by Bloom is currently the theoretical model for instruction in the FAA's Aviation Instructor's Handbook (FAA, 2017). For purposes of this study, the original Bloom's taxonomy is an excellent model for understanding how pilots think and learn to fly.

Bloom's Taxonomy and Flight Training

Cognition is concerned with the process of thinking and learning. Knowledge, perception, problem-solving, decision-making, awareness, and intellect are all critical elements of flight training. A subdomain of cognitive theory is constructivism, which describes learners as active participants in constructing knowledge and skills. Constructivist theory describes the process by which learners actively construct knowledge and skills based on prior knowledge and their experiences (FAA, 2017). As flight students apply higher order thinking skills (HOTS) during flight training, they are continuously constructing knowledge and skills. The researcher proposes that flight training is similar to Bloom's Taxonomy in that flight training builds upon each phase of pilot training and requires increased levels of complexity and performance as depicted in Figure 4.

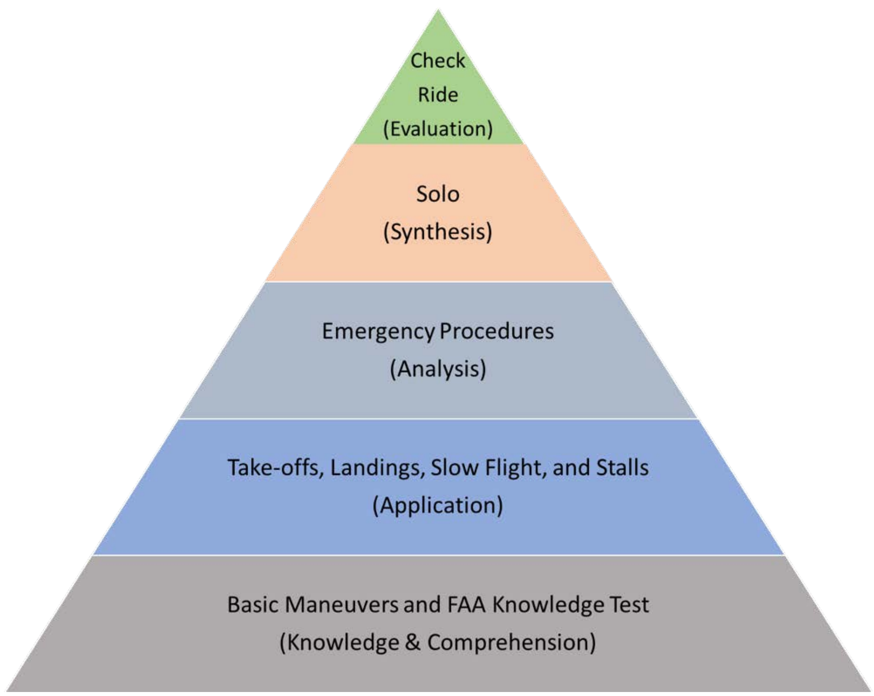

Figure 4: Bloom's Model and Pilot Training Levels (Adapted from FAA, 2017, p. 6-3).

Each level of the modified Bloom model for pilots depicted above builds on the previous levels, just as with Bloom's original model. In level one of the pilot model, the student pilot typically begins flight training by acquiring knowledge and comprehension of basic aircraft systems, rules, and regulations. This phase typically occurs during classroom instruction. Private pilots are required to take a written test and pass with a 70% or higher competence rating as one part of obtaining a pilot's license (FAA, 2018). Level two of the pilot model, application, requires considerable flight training time to learn and practice take-offs, in-flight maneuvers, and landings in both flight simulators and actual aircraft. In level three of the pilot model, analysis, emergency procedures require rapid analysis since the student has to first recognize the emergency and then properly address the emergency (e.g., engine failure) to ensure safe recovery in the aircraft. Level four of the pilot model, synthesis, occurs as the student conducts his or her first solo flight without an instructor. Level five of the pilot model concludes the pilot hierarchy with an FAA designated examiner check-ride evaluation. One can see that the modified Bloom model for pilots developed by the researcher provides a strong conceptual framework for discussion of pilot training.

High-level thinking skills are critical to ensuring good aeronautical decision-making. In flight training, HOTS are normally taught from simple to complex in the context of problem-based learning that emphasizes real-world problems. Scenario-based training is an excellent method of analytical problem-solving instruction. Scenario-based training can be utilized in both simulator and aircraft training to facilitate learning, perception, and skill transfer (FAA, 2017). Problem-based learning and scenario-based training require the pilot to use HOTS to solve real-world problems and hopefully, to transfer those skills to any future problems they may encounter. Therefore, by using scenario-based training coupled with higher-order thinking skills requiring the sense of sight, sound, and motor skills, a flight student should learn at a high degree of cognitive competence. However, a student's first solo and check-ride is only the beginning of a pilot's career. Advanced instruction in simulators and in complex aircraft continues throughout the majority of pilot flight training.

Bloom's Taxonomy of Cognitive Tasks (1956) provides the conceptual framework of this study of pilot learning and performance. The discussion that follows summarizes Sousa's (2017) seminal work in memory and learning. Sousa (2017) described critical components of memory and learning which are important when investigating flight simulation training and actual aircraft training.

Pilot Training

Working memory and long-term memory are critical when flying an aircraft, especially when a pilot is required to handle an in-flight emergency (Hoke et al., 2017). Sousa (2017) noted that working memory occurs in the brain's frontal lobes and captures a person's attention and focus. Cowan (2010) reported that researchers find it difficult to assign categories of working memory since it is affected by compounding variables such as interest, mental time delays, and distractions (Cowan, 2010). Learners who are more interested in a subject will be motivated to spend hours reading, working computations, and thinking critically about the subject. "An adult can normally process an item in working memory for approximately 10 to 20 minutes before reaching mental fatigue or boredom" (Sousa, 2017, p. 54).

Long-term memory is equally important to learning. "Long-term memory deals with the process of storing and retrieving information, while long-term storage refers to the areas in the brain where the memories are kept" (Sousa, 2017, p. 58). Long-term memory relates to the important ability to retain new information in order to apply it and to make connections to new tasks or information. Retention of information is important to ensure that learners can recall information at a later point in time. Neurological studies reveal that the greatest loss of new information occurs within the first 18 to 24 hours after exposure (Sousa, 2017). Therefore, if a learner has trouble recalling information after 24 hours, the information was probably not stored permanently in long-term memory.

Flight Simulation Training and Actual Aircraft Training

Bloom's Taxonomy, cognition, and memory play pivotal roles when comparing the value of flight simulation training and actual aircraft training. *Procedural memory* deals with the coordination of both motor and cognitive skills such as flying an airplane or driving a car (Sousa, 2017). *Procedural memory* is especially applicable to flying an airplane; the practice of flying skills is continuous, and procedural memories become efficient and can be performed with little conscious thought. When a flight student first begins learning to fly an airplane, he or she is cognizant of airspeed, maneuvering, and radio calls. However, when flying the airplane becomes more routine, skills are stored in procedural memory and become more automatic or reflexive (Sousa, 2017).

Loss of control during flight is one of the most common fatal mistakes in general aviation (Koglbauer, 2016). The procedure for recovering from aircraft loss of control normally relies on a series of memory items the pilot must perform during a specific, accelerated period of time. Attitude in an aircraft informs a pilot of the orientation of the aircraft relative to the Earth's horizon. "Safe recovery from an unusual attitude requires both declarative and procedural knowledge" (Koglbauer, 2016, p. 358).

Koglbauer (2016) conducted a study that researched the effects of simulator training on certified pilots' procedural memory. The flight simulator was a fixed computer-based trainer with two seats, controls, a throttle quadrant, and instrument panels. Thirty-one certified pilots with fewer than 450 hours of flight time participated in the study. The pilots were randomly assigned to a simulator training group (n = 17) and a control group (n = 14). All the participants participated in a written and an oral briefing from a certified flight instructor on the procedures for recovering from unusual attitudes. Additionally, each pilot observed a flight demonstration by a flight instructor in a Pitts S-2B aircraft on ways to recover from unusual attitudes. After the S-2B flight demonstrations were completed, the simulator training group practiced the recovery procedure nine times in the simulator. Once the simulator training was completed, both the training group and the control group conducted a post-test flight with the flight instructor in the Pitts S-2B aircraft.

The results of the Koglbauer (2016) experiment indicated that the simulator training group performed the practiced attitude recovery significantly better than the control group ($p < .001$), which did not perform the recovery in the simulator. Additionally, the simulator training group completed the practiced recovery procedure significantly more quickly than the control group ($p < .01$). When introduced to a new maneuver by the flight instructor, the simulator training group performed recovery of the new maneuver better than the control group, but the differences were not statistically significant ($p < .09$). The researchers found positive correlations between the practiced and new maneuver ($p < .05$) and between the task completion time of the practiced and new maneuver ($p < .001$). Koglbauer's (2016) study revealed that when pilots used a simulator to practice a maneuver related to unusual attitudes, a combination of procedural and declarative memory was improved; but no change was observed when a different maneuver was introduced.

This study implies that transfer of training involves more than procedural and declarative memory and points to the consistent need for practice in a variety of unique situations when flying. Memory is vital in pilot training, but to exercise memory, pilots need to rehearse skills and practice either in simulation or in an actual aircraft in order to retain pilot knowledge.

Rehearsal and Practice

An accident analysis and prevention study was conducted by Molesworth, Bennett, and Kehoe (2010) to examine ways that skill rehearsal could mitigate pilot failure. A personal computer (PC) was used to simulate a single-engine fixed wing Cessna 172 aircraft. The PC flight simulator consisted of precision flight controls and rudder pedals and utilized X-Plane 6.21 simulation software. The study included 30 pilots from a local flight training institution with a mean flight time of 152 hours. Nine participants held recreational pilot licenses, 11 held private pilot licenses, nine had commercial pilot licenses, and one had an Air Transport Pilot (ATP) license. Participants were randomly assigned to one of three groups: a self-explanation group (SE), a relapse prevention group (RP), and a control group (C); each group had ten members. Following one simulation flight, the SE group completed a self-explanation questionnaire, the RP group completed a relapse-prevention questionnaire, and the control group conducted the simulation flight without any debriefing. Each participant in the study was required to maintain altitude at 500 feet above ground level (AGL) over a 20-minute period based on a variety of general aviation low-flight training-based scenarios in the PC simulator. The control group completed the simulation and was dismissed; the SE and RP groups completed written self-explanation and reflection questionnaires after the simulation. The self-explanation questionnaires required pilots to explain their actions after the simulated flight and actions they would take in future flights. The reflection questionnaires required pilots to self-reflect on behavior that might have led to unsafe altitudes. The researchers found that group SE tended to stay above 500 feet, group RP tended to fly below 500 feet, and group C demonstrated poor compliance with the 500 feet minimum. However, differences between the three groups were not statistically significant ($F < 1$). After a single simulated flight, the self-explanation reflection group demonstrated greater ability to stay above the 500 feet AGL safety rule using a flight simulator. The authors suggest that pilots may increase learning when reflecting on and recalling experiences in rehearsal. Additionally, rehearsal reflection may also help sustain deeper mental processing and later retrieval of memory when future application warrants (Molesworth et al., 2010).

Rehearsal and practice are sometimes considered the same thing. However, practice is different from rehearsal. Practice implies that a learner repeats a skill over a period of time whereas rehearsal implies that the learner reprocesses new information in an attempt to determine sense and meaning (Sousa, 2017). "When first learning a skill, attention and awareness are obviously required" (Sousa, 2017, p. 108). Eventually, procedural memory takes over, and the performance of the skill becomes almost automatic (Hirano, Kubota, Tanabe, Koizume, & Funase, 2015). In flight training, a student typically practices a certain task repeatedly over a period of time (e.g., landings). However, if practice is halted for any reason, the brain may re-assign neurons that are no longer being used to other tasks, and the student's skills may decline (Sousa, 2017), which is the reason the FAA requires licensed pilots to continuously train to refine and maintain their skills. For example, a private pilot must complete three take-offs and landings every 90 days to stay current and be qualified to carry passengers (FAA, 2018).

Guided practice during flight training is similar to a flight instructor's presence during take-offs, landings, and maneuvers before letting the flight student solo and practice solo. Unlearning and relearning skills can be challenging and cause frustration or loss of motivation to complete training. Sousa (2017) suggests that effective practice first starts with guided practice, followed by independent practice, and eventually distributed practice, helping the learner to learn from errors and to continuously refine skills. Pilot training, whether in simulations or actual aircraft, utilize similar teaching patterns when moving from instructor-driven practice and rehearsal to the achievement of sufficient competency to solo and eventually fly in combat. Guided practice, independent practice, and distributed practice can be accomplished in either the simulator or actual aircraft.

Transfer of Training

Sousa (2017) reports that transfer of knowledge occurs in a two-part process: transfer *during* learning and transfer *of* learning. *Transfer during learning* describes past learning and its effects on new learning, whereas *transfer of learning* describes the ways that the learner applies the new learning to new situations in the future. There are two common types of transfer: positive transfer and negative transfer. *Positive transfer* occurs when past learning assists the student with new learning (Sousa, 2017). *Negative transfer* occurs when past learning differs from new learning; the result is often confusion or errors and is referred to as negative training (Sousa, 2017). For example, if a pilot learns a specific procedure in a simulator that is a little different from a real airplane, he or she could develop a bad habit or have trouble re-learning a specific skill.

Pfeiffer, Horey, and Butrimas (1991) conducted a study using 34 undergraduate flight students at Training Wing TWO in Kingsville, Texas. The trainees were randomly assigned to one of three training levels (2, 4, or 8 simulator flights) and to one of two aircraft flights focused on visual or basic instruments. All 34 participants conducted ten flights in which two were conducted in the T-2C aircraft.

Flight training was initially conducted in the T-2C flight simulator followed by continuation training in the T-2C aircraft. Performance measures were recorded by instructors during the simulator flights and two actual aircraft flights. The performance measures consisted of deviations from assigned airspeed, altitude, heading, and turning maneuvers (Pfeiffer et al., 1991). Data were collected and segregated into four groups: simulator training level, flight rule, maneuver, and sequence. The data were summed across the groups, and an analysis of variance (ANOVA) was conducted. The results indicated significant transfer of training from the simulator to the aircraft ($p = .007$). The researchers also found that pilots demonstrated positive transfer of skills after training in a flight simulator in order to transition from a T-34C, a single engine turbo-prop aircraft to a T-2C turbojet aircraft.

Many factors affect the learning transfer process, such as the context and degree of original learning, similarity of new learning to prior learning, critical attributes of the task, and association or connections between prior learning and new learning (Sousa, 2017). “The quality of transfer that occurs during new learning is largely dependent on the quality of the original meaning” (p. 160). Similarity describes the cognitive processes by which the skills learned in one environment can be transferred to other similar environments (Sousa, 2017). Commercial jet pilots are first trained in flight simulators before they fly the actual commercial airplane. The training and learning that occur in the simulator, a replica of the actual plane, should, theoretically, transfer to the actual flying environment.

This review of literature on the theoretical and conceptual models of learning, as well as the previous research described, all point to the vital role of both flight simulation training and actual aircraft training in pilot training. However, the effectiveness of flight simulation training and its relationship to actual aircraft training is an essential subject that demands further research.

Overview of Training Research

The research on training humans encompasses a large body of disciplines including industrial/organizational psychology, military psychology, human factors, and more (Bell, Tannenbaum, Ford, & Noe, 2017). However, the current study was primarily concerned with pilot training research. In the early years of pilot training research, (1917-1958) most studies focused on training efficiency, training success in the military (primarily aviation-related), and human relations (Bell et al., 2017). The first known article on pilot training was originally published by Geissler (1918) shortly after the start of World War I (Bell et al., 2017). Geissler (1918) discussed the skills needed to pilot a glider and other training tasks. Kellogg (1946) collected data using a pilot-response recorder to investigate ways the data could be used to improve pilot skills. He found that recordings of airplane control surface movements could demonstrate pilot progress in the development of flying skills, specifically landings.

Fiske (1947) conducted a study of flight training candidates' ground school failures to ascertain whether the information could be useful for predicting future flight failures. Three selection tests were used by the Navy prior to World War II: the *Personnel Test,* the *Mechanical Comprehension Test,* and the *Biographical Inventory*. The study included 6,247 Navy flight training cadets from 1941-1943. The study found that the *Personnel Test* was useful in helping to predict ground school failures. The *Mechanical Test* was significantly related to pilot training and flight failures. In the late 1950s and early 1960s, transfer of learning research became prominent. Pilot training research in the 1980s and 1990s focused primarily on learners, learners' views of training, and the ways learning was tied to experience.

A study conducted by Smith-Jentsch, Jentsch, Payne, and Salas (1996) revealed that pilots who had been pressured to fly in bad weather were more motivated to learn about the adverse effects of flying in bad weather. Based on a sample of 32 private pilots, the study found a linear relationship between negative flight events and pilot assertiveness after one week of training. In other words, pilots were more assertive about their arguing not to fly in bad weather after receiving flight training.

The 2000s saw an increase in published pilot training articles and research based on new evaluation tools, training effectiveness, and new training models such as eLearning (Bell et al., 2017). Referring to pilot training, (Bell et al., 2017, p. 315) wrote, “We also need to continue to evolve how we think about and measure learning.”

Pilot Training Effectiveness

Kirkpatrick’s (1959) seminal study of training in manufacturing and business organizations was one of the first published articles to measure training effectiveness. He described four categories of measurement of training effectiveness. The first category, *reactions,* pertains to the ways trainees perceived a particular training program. This type of measurement might involve course evaluations, surveys, focus group discussions, and other descriptive measures of participants’ perceived effectiveness of training. Kirkpatrick’s second category was *learning,* which described the skills, facts, and principles understood by the trainees; this type of learning could be measured by knowledge and comprehension tests. The third category was *behavior,* measured by participants’ application and use of learned principles and behaviors on the job. The fourth and final category was *results* as measured by participants’ achievement and implementation of the desired training goals over time to increase production and to improve trainee morale. Kirkpatrick argued that all training programs are measured by all four of these categories, and his work has been largely accepted and applied in many studies and organizations today.

Some have argued that Kirkpatrick's (1959) model for training effectiveness was based on large assumptions. Alliger and Janak (1989) stated that the categories are linked, vague in description, and are positively correlated. For example, students may respond positively to a training program, but they may not be able to apply the concepts on the job (Alliger & Janak, 1989). In addition, training effectiveness can be affected by a number of non-technical factors, such as participants' self-efficacy, task-related attitudes, expectations about training, training fulfillment, and motivation (Cannon-Bowers, Tannenbaum, & Mathieu, 1995). Organizational culture, policies, and instructional methods can all shape how trainees perceive the training (Cannon-Bowers et al., 1995).

Povenmire and Roscoe (1972) conducted an experiment on the training effectiveness of a flight simulator and an actual aircraft. The goal of the experiment was to determine whether a positive relationship existed between ground-based training in the Link G AT-1 flight simulator and possible cost savings by using a Piper Cherokee aircraft for actual flight training. The study consisted of 65 flight students randomly assigned to four experimental groups. The researchers found that the effectiveness of a ground-based flight simulation trainer depended on the fidelity of the simulation, reliability of the simulator, and the training of the flight instructor to use the simulator. The results revealed slight cost savings, but due to the time period (1972), the simulator cost per hour ($16) compared to the Piper Cherokee's cost per hour ($22) was not substantial.

Sandia National Laboratories developed a performance-based assessment tool in 2002 called the *Automated Expert Modeling and Student Evaluation System* (AEMASE) to measure pilot training effectiveness. AEMASE simulates human behavior using machine-learning to build an expert model of military tactics and then comparing student behaviors on flight tasks against the expert model to establish a score. Stevens-Adams, Basilico, Abbott, Gieseler, and Forsythe (2010) conducted a study to determine whether the AEMASE system provided useful feedback to US Navy E-2 Hawkeye radar operators in order to enhance training effectiveness. Twenty-two participants who met the requirements for an entry-level US Navy E-2 Hawkeye radar operator were split into two groups: a control group (n=12) and a debrief group (n=10). The study used two of the participants as E-2 Hawkeye subject matter experts. Both the control group and debrief group completed five simulation-based training sessions provided by an E-2 instructor who taught the basic operations of the E-2 radar system in the simulator. The debrief group was given detailed performance feedback by the AEMASE debrief tool including real-time, verbal feedback using the timeline and graphical radar map depictions provided by the AEMASE tool. The control group received real-time, verbal feedback by the instructors on their performance deficiencies in radar operations. Finally, both the debrief and control groups were given two testing sessions in which they completed five difficult simulations without the aid of an instructor. The researchers found that participants in the debrief group identified radar threats significantly more quickly than the control group ($t = 2.03$, $p < 0.05$). Participants in the debrief group also identified neutral radar threats significantly more quickly than the control group ($t = 1.87$, $p < 0.05$). These results indicated that the debrief group that used the AEMASE tool performed significantly better on certain tasks than the control group, which had received only verbal feedback. Though the study assessed E-2 radar operators and not pilots, the results are important to demonstrate ways that simulation and debriefs can enhance training effectiveness.

The US Navy continuously assesses the need to increase military readiness and to identify training solutions that can increase training effectiveness. Betts, McCauley, and Walwanis (2010) administered a questionnaire to 77 Navy operational F/A-18 pilots to identify the types of deployable training systems (small, portable systems that can be transported, such as simulators) that would maintain flying skills while deployed. The study attempted to determine F/A-18 skills that required the most training and F/A-18 skills that were the most perishable during deployment. The questionnaire contained four sections: 1) mission tasks ratings, which asked participants to rate each flyable mission task on a five-point scale, 2) free responses, which allowed participants to identify their top ten training requirements while deployed, 3) deployable system requirement ratings, which asked participants to rate each training system's capability on a five-point scale based on the need for training, and 4) demographic information on rank, time in service, and total flight hours of each participant.

The data from the questionnaire were compiled for all 77 respondents, and a mean score and standard deviation of the *Skill Decay Index* (SDI) were computed for each of 78 identified mission tasks (Betts et al., 2010). The skill decay analysis was based on two types of factors: 1) factors that were task-related and 2) factors that were user-related (Swezey, Owens, Bergondy, & Salas, 1998). The skill decay index was computed by summing all scores for a particular task; scores ranged from 3 to 9. The composite SDI score for all 78 mission tasks was 6.18 with a standard deviation of 1.15. The participants ranked Close Air Support (CAS) as the most critical mission task requiring training (*Mean* SDI = 7.96, *SD* = 1.27). CAS consists of providing live aircraft weapon support to ground troops or other top-ranked mission tasks including air-to-air missions (combat in the air with other enemy aircraft) and air-to-ground missions (destruction of ground-based targets such as enemy radar). The CAS rankings in this study were not surprising since fewer training opportunities are available to perform CAS, air-to-air, and air-to-ground missions unless a real threat emerges while operationally deployed (Betts et al., 2010). According to Betts et al. (2010), a large percentage of respondents in the free-response survey items identified the need for a complex simulator training system located on an aircraft carrier rather than on a deployable laptop-based training system. The top three deployable training system requirements identified by the participants were accurate cockpit displays, aircraft system replication, and high-fidelity controls capabilities. These responses suggest that “the training system has to perform like the jet in all respects in order to be effective, or even used” (p. 7). Betts et al. (2010) concluded that a deployable simulator training system would not ameliorate skill decay. Deployed pilots cited the need for more flight hours as the reason that a deployable training system would not meet all their required training needs. These responses revealed that pilots believed that actual aircraft training time was needed in conjunction with simulation training time. The study’s authors recommended that the deployable training system be used not to replace flight hours, but to be used in conjunction with aircraft hours to make actual flight time more effective.

In 2012, the US Army conducted a holistic analysis of the Army’s doctrinal framework by reviewing data on pilot transfer of training, cost-benefit data, and the potential benefits of increasing the use of simulators for flight training. The analysis included only active duty aviators and did not address aviators in the National Guard or Army Reserve (Blow, 2012). The Army identified actual aircraft training as the foundational method for effective pilot training, citing that simulators lacked the fidelity or proficiency to replace aircraft training (Blow, 2012). However, the Army also concluded that despite the simulators’ low fidelity, the simulators could produce quality training if used correctly. Simulators could specifically be used to train primary tasks and instruments and to reduce overall training costs (Blow, 2012).

Flight Simulator Effectiveness

US Navy training has traditionally focused on live training events; however, ongoing budget constraints have served to promote the use of simulators to train pilots. (Schank et al., 2002). In 2002, the Navy asked the RAND National Defense Research Institute to study the ways that a mix of classroom instruction, simulation training, and aircraft training events could increase pilot training effectiveness. The RAND study conducted by Schank et al. (2002) focused on two airborne training programs within the Navy's F/A-18 and the P-3C squadrons. Additionally, RAND also conducted a trade-off analysis between live and simulated training events while analyzing flying hours and simulator use for US Navy units. However, the data were based on best estimates and data that were readily available. The study found that F/A-18 simulators made only a modest contribution to flight training due to poor accessibility by pilots and low fidelity; however, P-3C simulators were more widely used and accepted by the P-3C pilot community. The P-3C simulators were used more often than the F/A-18 simulators by pilots, who judged that P-3C simulators replicated the flight environment better than the F/A-18 simulators.

The Navy employs four types of F/A-18 simulators, although some simulators are in the process of being replaced or are currently in the process of hardware and software upgrades (Schank et al., 2002). The RAND study found that the F/A-18 simulators were used by the fleet training schools 50% to 60% of the time; however, fleet operational units (deployed units) used the simulators 20% of time, and they were unused 20% of the time. The RAND study also conducted interviews and determined that pilots frequently do not use the simulators due to a lack of realistic mission profiles, lack of fidelity between the simulation and actual flight time, and lack of availability. While training schools used the simulators 50 to 60 percent of the time;

once deployed, pilots used them very little if at all. The results of this study are intriguing. In 2002, pilots recognized the limitations and strengths of different types of simulators and made their opinions known, both verbally and in their non-use of the simulators.

A recent evaluation study by Koglbauer, Riesel, and Braunstingl (2016) was conducted on the effects of combining actual aircraft and simulator training on student pilots' skill acquisition. The evaluation consisted of 61 general aviation flight students with zero flight hours and approximately 40 hours of classroom training. The evaluation employed a pre-test and post-test design using the repeated measures *t*-test to compare simulator scores and scores on aircraft flight tests. The alpha level was set at 0.05, and the *t*-test of dependent samples and Pearson's coefficient of correlation were used to analyze the collected data. The results indicated that the students' aircraft flight performance scores were significantly higher on the post-test ($p < .01$) after using the simulator. Also, a significant positive correlation was found between instructor grades in the simulator and the aircraft flight post-test ($p < .01$). The study indicated that a combination of actual aircraft training and simulator training both have positive effects on beginning pilots during actual flight.

Trade-Offs Between Actual and Simulated Training

The RAND study in 2002 reported that very few experimental studies had been conducted comparing the benefits of actual and simulated aircraft training; however, Schank et al. (2002) found non-experimental studies identifying the possibility of substituting simulator training for actual aircraft training. One such study conducted by Roof (1996) investigated F/A-18 pilot qualification events to determine whether pilot skills could be developed effectively in the F/A-18 simulator. Roof reviewed data from government publications, professional materials, previous theses, books, and articles. Additionally, Roof (1996) conducted interviews with Naval aviation program officers and readiness officers in Naval Aviation and Manpower Training Offices. Based on the interviews and the Training and Readiness Matrix (TRM), the author created a F/A-18 *must-fly* criterion and cross-referenced the remaining training events to the F/A-18 simulator. The *must-fly* criterion included: qualifications that require a substantial amount of maneuvering, qualifications that require extensive visual cues, and qualifications that require concentrated communications with other aircraft. Roof (1996) concluded that substantial cost savings could be realized by moving certain tasks, excluding the *must-fly* tasks, from live aircraft training to a simulator with no substantial degradation in training or safety. The cost savings were based on F/A-18 aircraft cost per hour ($2,973) and F/A-18 simulator cost per hour ($432), as reported in Schank et al. (2002).

Based on the RAND study (2002), four major trade-off recommendations were noted: 1) the military could save costs by using simulation hours for flight hours, 2) an increase in number of flight simulator hours to actual aircraft hours could possibly help increase pilot proficiency, but little empirical data exist to support this change, 3) the Navy could modify the Navy's training events to achieve the same level of readiness with an increase in simulation hours, and 4) the military could increase simulator fidelity and aircraft availability. According to Schank et al. (2002), simulator hours have to be encouraged as part of a pilot's culture. "Time in a pilot's day, simulator fidelity and availability, the availability of aircraft, and pilot experience can all be treated as variables and changed to explore different trade-offs" (Schank et al., 2002, p. 59). The RAND study concluded that the use of simulators could increase, and a trade-off could be made by moving live training events to a simulator, depending on the availability of the simulators. However, the authors specified necessary improvements to simulations including: additional funding to improve simulator fidelity, greater simulator participation by pilots, and the use of simulators as a complement to live training, but not necessarily as a substitute for live training. “Training proficiency depends, among other things, on task characteristics and complexity” (Schank et al., 2002, p. 51). Carretta and Dunlap (1998) conducted a study that found that pilots with 40 to 60 simulator sessions demonstrated significantly better landing skills than pilots with 20 simulator sessions. The research conducted by RAND also suggested that independent of simulation accuracy or fidelity, some training events are best flown in an aircraft, such as safety-of-flight events and events requiring essential physiological cues (e.g., high G maneuvering).

Since few experimental studies exist on the effectiveness of simulator training compared to actual aircraft training, the RAND study was based largely on aviators’ experiences and reports. Analysis of descriptive and qualitative research data reached the following conclusions regarding the use of flight simulators for pilot training: (a) simulation is effective in introduction, practice, procedural training, and for rehearsal; (b) live training may be needed to learn specific motor skills, (c) simulators are not a substitute for actual flight time, (d) pilot experience is an important factor in the value of simulation training or effectiveness, (e) aircraft type and complexity influence the value and effectiveness of simulation training, and (f) negative training should be avoided since simulators can teach wrong responses and possibly give a false sense of accomplishment.

The RAND study’s overall recommendations to the U.S Navy stated that the Navy must measure pilot readiness more accurately, identify the future combination and balance of training (classroom, simulator, and aircraft), determine whether higher levels of fidelity are possible in simulators and whether they are cost effective; identify both physical and psychological limits of live and simulated training; and purchase more and better simulators if the Navy determines that simulators provide effective training.

Summary

The literature review discussed the theoretical underpinnings of this study and the literature on flight training to establish the rationale for this study's purpose of comparing flight simulation time versus actual aircraft training time and the overall impact on pilot performance. Chapter 3 presents the research methodology used to analyze actual aircraft hours and flight simulation hours and their relationships to the Naval Standard Score of USN and USMC intermediate and advanced student pilots.

III. METHODOLOGY

This chapter describes the methods used in this study to compare T-45C Goshawk aircraft training hours and flight simulation training hours and their relationships to the Naval Standard Score (NSS) of USN and USMC intermediate and advanced flight students. The study was non-experimental, correlational, causal-comparative research using archival data from the Chief of Naval Air Training (CNATRA) Training Information System (TIMS) database. CNATRA aircraft hours, flight simulation hours, and NSS scores of intermediate and advanced flight students from 2015 to 2017 were analyzed and compared. The research study was approved as exempt by the Southeastern University (SEU) Institutional Review Board (IRB) and the US Navy IRB.

Participants

The study's sample was purposive and obtained from CNATRA's TIMS database from years 2015 to 2017. The dataset was purposive since the data were selected for only intermediate and advanced flight students who trained with the T-45C Goshawk and T-45C Operational Flight Trainer (OFT). Additionally, the years 2015 to 2017 were chosen to ensure reliability and validity by limiting changes over time in simulation upgrades or aircraft technology. Participants included all USN and USMC intermediate and advanced T-45C Goshawk flight student completers from 2015 to 2017 from Training Wings ONE and TWO.

The sample size consisted of 358 intermediate flight students and 334 advanced flight students from Training Wings ONE and TWO. Flight students are required to complete a minimum number of aircraft training hours and flight simulation hours, but students can request more unproctored time in the simulator; as a result, simulator times are different for each student. The focus of the study was designed to compare actual aircraft training hours and flight simulation training hours, and their relationships to the Naval Standard Score (NSS) of intermediate and advanced USN and USMC flight students. The independent variables were actual aircraft hours and flight simulation hours. The independent variables were continuous, ratio-level data. The dependent variable was the NSS score, which is was normalized by the Navy as interval-level data with a NSS range of 20 to 80, mean of 50, and a standard deviation of 10.

The researcher did not analyze any demographic data such as gender, military rank, or age. The study examined only USN and USMC actual aircraft hours, flight simulation hours, and the NSS for years 2015 to 2017. The study investigated the relationships of one aircraft, the T-45C and flight simulator, the T-45C OFT, in order to control key variables and to help ensure comparability of data.

Instrumentation

The independent variables were actual aircraft hours and flight simulation hours. Actual aircraft hours were flight hours completed in the T-45C Goshawk aircraft as recorded in the CNATRA TIMS database. Actual aircraft hours were flown in the T-45C Goshawk, a highly maneuverable turbofan jet military training aircraft built by the Boeing Company (Boeing, 2018). The T-45C consists of a digital cockpit with integrated navigational displays, weapon delivery systems, and communication systems. Flight simulation hours were gathered from the T-45C Goshawk OFT as recorded in the CNATRA TIMS database. The T-45C OFT is a high-fidelity instrument and visual flight simulator designed as a dome shell simulator with a digital cockpit, outside visual displays, and instructor station (Boeing, 2018).

The dependent variable in this study, the NSS, was used to measure pilot competencies and skills for both the USN and USMC flight students. CNATRA considers the NSS score to be a valid method of assessing pilot competency for the Navy's flight training program (Naval Air Training Command, 2014). The NSS is calculated separately for each flight student who completes each phase of training (primary, intermediate, and advanced). The NSS is a normalized, cumulative score ranging from 20 to 80 and is based on training events' scores. The mean NSS is 50 with a standard deviation of 10. During all three levels of training, a NSS score greater than 50 is considered by the Navy to be above average, and a score of 60 would imply one standard deviation above average. Navy flight students completing primary training with an NSS score less than 50 are not eligible to pilot strike aircraft (e.g., F/A 18); Marine Corps flight students with an NSS less than 52 are not eligible to pilot Strike aircraft (Naval Air Training Command, 2014). This study focused on training times of flight students in the intermediate and advanced phases of training.

The Naval Standard Score (NSS) is calculated by using the equation $NSS = 50 + 10 * (PAS - MPAS/SDPAS)$. The Phase Aggregate Score (PAS) is a comparative ranking based on flight students' performance scores on a group of training events compared against the performance scores of a previous population of pilot training completers for the same set of training events (Naval Air Training Command, 2014). The PAS score of previous completers is based on 200 previous students from the same squadron for primary training and the previous 60 students for intermediate and advanced training.

Procedures

The following research questions and null hypotheses were posed in this study:

Q1: Which is the best predictor of the NSS scores of intermediate pilot trainees: flight simulation time or actual aircraft time?

$H_0$1: Flight simulation time and actual aircraft time are not significant predictors of NSS scores of intermediate pilot trainees.

Q2: Which is the best predictor of the NSS scores of advanced pilot trainees: flight simulation time or actual aircraft time?

$H_0$2: Flight simulation time and actual aircraft time are not significant predictors of NSS scores of advanced pilot trainees.

Q3: Are there any significant differences between the intermediate and advanced pilot trainees' flight simulation time and actual aircraft time?

$H_0$3: There is no significant difference between mean flight simulation time and mean actual aircraft time of intermediate and advanced pilot trainees.

Data were collected from the CNATRA TIMS database and exported into a Microsoft Office Excel file. Flight student data were then imported into *Statistical Package for the Social Sciences Version 24 (SPSS)* for analysis purposes.

Data Analysis

Descriptive statistical techniques were used to analyze the archival dataset. Specifically, missing data, essential demographic variables, and data array normality tests was conducted for reporting purposes. The mean was calculated for flight simulation hours, actual aircraft hours, and the NSS scores for both intermediate and advanced flight training students. Additionally, Cohen's *d* and Cohen's *q* were used to evaluate the magnitude of effect sizes. The alpha level of .05 was utilized as the threshold value for determining statistical significance.

In advance of comparative analyses requiring the assumption of normality, the study's data arrays were assessed for normality using the Shapiro-Wilk test. To address the first and second research questions, multiple linear regression was used to determine the significant predictor(s) of the NSS scores. The independent variables were flight simulation time and actual aircraft time; the dependent variables were the NSS scores of intermediate pilot trainees and of advanced pilot trainees. Using the Fisher *r* to *z* transformation test statistic, the difference in correlations between simulation time and aircraft time and NSS was evaluated for statistical significance. Cohen's *q* test statistic was used to evaluate the effect size of correlations between simulation time and aircraft time. The alpha level of .05 was utilized as the threshold value for determining statistical significance.

In light of the non-normal distribution of the data related to the comparison in Research Question 3, the Mann-Whitney U test was used to evaluate the statistical significance of the difference between mean ranks for training type (flight simulation time versus actual aircraft time) and training phase (intermediate and advanced). Cohen's *d* statistic was used to evaluate the magnitude of difference of mean ranking (effect size) in the respective comparisons. The alpha level of .05 was utilized as the threshold value for determining statistical significance.

Ancillary analyses were also conducted using Chi-Square to compare the intermediate and advanced training groups' mean scores on the NSS to determine whether significant differences existed. Scores on the NSS were disaggregated into three different NSS sets (SS $\leq$ 50, NSS +1 Standard Deviation (*SD*), and NSS + 2 *SD*) and then compared based on trainee group. The alpha level of .05 was utilized as the threshold value for determining statistical significance of differences.

Summary

The purpose of this study was to investigate the relationships between flight simulation training times, actual aircraft training times, and intermediate and advanced jet pilot competence as measured by the Naval Standard Score (NSS). The study was non-experimental, correlational, causal-comparative research using archival data from the CNATRA TIMS database. This study also examined ways that flight simulation hours and actual aircraft training hours related to the NSS score. The methodology included multiple linear regression, the Fisher *r* to *z* transformation test, and the Mann-Whitney U test. The Cohen's *q* test statistic was used to evaluate the effects size of correlations between flight simulation time and actual aircraft time. The Cohen's *d* statistic was used to evaluate the magnitude of difference of mean ranking (effect size) in the respective comparisons. The alpha level of .05 was used as the threshold for determining statistical significance. The results of the research methods and analyses are presented in Chapter IV.

IV. RESULTS

The purpose of this study was to investigate the relationships between aircraft training, flight simulation training, and intermediate and advanced jet pilot competence as measured by the Naval Standard Score. Flight simulation time and actual aircraft time were used to determine the best predictor(s) of the NSS score of intermediate and advanced pilot trainees. Data were obtained from the Chief of Naval Air Training (CNATRA) Training Information System (TIMS) database for years 2015 to 2017. Differences between intermediate and advanced pilot trainees based on flight simulation time and actual aircraft time were examined and compared. Ancillary analyses were also conducted to further examine the relationships between flight simulation time, actual aircraft time, and the NSS scores.

Preliminary Analyses

Prior to formally addressing the research questions, preliminary analyses of the study's data were conducted. Specifically, missing data analysis, data array normality analyses, and descriptive analyses of study variables were conducted.

Frequency and percentage of missing data was analyzed using descriptive statistical techniques. With regard to the three essential data arrays (simulation hours, flight hours, and NSS scores), a total of nine data points of a possible 2,073 (0.43%) were identified as missing. In light of the minimal degree of missing data (< 1.0%), the researcher determined that imputation of missing data points within the essential arrays was not necessary.

In advance of comparative analyses requiring the assumption of normality, the study's data arrays were assessed for normality/relative normality using the Shapiro-Wilk test. Results of the Shapiro-Wilk test are depicted in Table 1.

Table 1

Shapiro-Wilk Test Results by Training Type, Training Group, and NSS score

Group	SIM Hours			Flight Hours			NSS		
	S-W	*df*	*p*	S-W	*df*	*p*	S-W	*df*	*P*
Intermediate *n* = 334	0.86	328	.000*	0.93	328	.000*	0.99	328	.003*
Advanced *n* = 358	0.86	357	.000*	0.83	357	.000*	0.99	357	.68[a]

[a] $p > .05$ *$p < .05$

As seen in Table 1, only one of the six essential data arrays evaluated, Advanced Group NSS Score*,* was found to be normally distributed ($p > .05$). As a result, analyses of non-normal data were required and are discussed in upcoming paragraphs.

Descriptive statistics were used to describe the study's participants and their scores on the independent and dependent variables. The results are depicted in Table 2.

Table 2

Mean and Standard Deviations on Intermediate and Advanced Flight Students Statistics

	SIM Hours			Flight Hours			NSS		
Group	*n*	Mean	*SD*	*n*	Mean	*SD*	*n*	Mean	*SD*
Intermediate	334	84.72	5.98	336	81.34	8.45	330	48.86	11.15
Advanced	358	24.92	3.24	358	85.64	8.94	357	50.00	10.66

Table 2 reveals that mean simulation hours for the intermediate group were considerably higher than simulation hours for the advanced group. However, mean flight hours for both training groups were more similar and the standard deviations were smaller than for simulation hours.

Analyses of Research Questions and Hypotheses

Three research questions and null hypotheses were posed in the study to address the stated research problem. The questions and hypotheses were evaluated analytically.

Research Question and Hypothesis 1

Q1: Which is the best predictor of the NSS scores of intermediate pilot trainees: flight simulation time or actual aircraft time?

$H_0$1: Flight simulation time and actual aircraft time are not significant predictors of NSS scores of intermediate pilot trainees.

Research Question 1 was predictive in nature, involving two independent covariates. A multiple linear regression was conducted to evaluate the predictive abilities of flight simulation training time and actual aircraft training time in the same model. Table 3 depicts the results of the statistical test for the intermediate group of pilot trainees

Table 3

Prediction of NSS by Training Type of Intermediate Group

Model	*β*	*Standard Error*	*Standardized β*
Intercept	100.26	9.39	
Simulation Training	-0.02	0.12	-.01
Aircraft Training	-0.61	0.12	-.32***

****p* < .001

The results indicated that actual aircraft training time was a strong, significant ($p < .001$) predictor of NSS scores of intermediate pilot trainees. Flight simulation training time was not a significant predictor of NSS scores of intermediate pilot trainees. Therefore, the null hypothesis ($H_0 1$) was rejected. The regression weight ($\beta = -0.61$) means that for every one unit increase in aircraft training time there is a -0.61 decrease in the NSS score.

The significant, predictive effect for actual aircraft training method may be considered medium (ES = .12) (Field, 2013). Using the Fisher *r* to *z* transformation test statistic, the difference in correlations between flight simulation time, actual aircraft time, and NSS scores was statistically significant ($z = -4.10$; $p < .001$). Using the Cohen *q* Test Statistic to evaluate the comparative relational effect between the two training methods, the magnitude of effect (favoring actual aircraft training) was considered medium (Cohen's $q = 0.32$) (Cohen, 1988).

Research Question and Hypothesis 2

Q2: Which is the best predictor of the NSS scores of advanced pilot trainees: flight simulation time or actual aircraft time?

$H_0$2: Flight simulation time and actual aircraft time are not significant predictors of NSS scores of advanced pilot trainees.

Research Question 2 was predictive in nature, involving two independent covariates. Multiple linear regression was used to evaluate the predictive abilities of both flight simulation training time and actual aircraft training time in the same model. Table 4 depicts the results of the statistical test.

Table 4

Prediction of NSS by Training Type of Advanced Group

Model	*β*	*Standard Error*	*Standardized β*
Intercept	63.32	6.15	
Simulation Training	0.13	0.23	.04
Aircraft Training	0.19	0.08	.15*

$^*p < .05$

Aircraft training time was a statistically significant predictor ($p < .05$) of NSS scores of advanced pilot trainees. The predictive effect for the aircraft training method was considered small (ES = .02) (Field, 2013). Using the Fisher *r* to *z* transformation test statistic, the difference in correlations between flight simulation time, actual aircraft time, and advanced pilots' NSS scores was not statistically significant ($z = 1.48$; $p =.07$). Using the Cohen *q* test statistic to evaluate the comparative relational effect between the two training methods, the magnitude of effect (favoring aircraft training) was considered small (Cohen's $q = 0.11$) (Cohen, 1988). The regression weight ($\beta = 0.19$) means that for every one unit increase in aircraft training time there was a 0.19 increase in the NSS score. Because the results indicated that actual flight time of advanced pilots was statistically significant, the null hypothesis ($H_0 2$) for Question 2 was rejected.

Research Question and Hypothesis 3

Q3: Are there any significant differences between the intermediate and advanced pilot trainees' flight simulation time and actual aircraft time?

$H_0$3: There is no significant difference between mean flight simulation time and mean actual aircraft time of intermediate and advanced pilot trainees.

In light of the non-normal distribution of the data arrays related to the comparison inherent in Research Question 3, a non-parametric test statistic alternative to the *t test* of independent means was utilized for analytical purposes. Specifically, the Mann-Whitney U test was used to determine whether significant differences existed between mean ranks of training type (flight simulation time and actual aircraft time) and training groups (intermediate and advanced). Cohen's *d* statistic was used to evaluate the relative magnitude of difference of mean ranks (effect size) in the respective comparisons. The results of the analyses are depicted in Table 5.

Table 5

Mann-Whitney U Test and Cohen's d of Training Type and Training Group

Comparison	*n*	Mean Rank	*z*	*p*	*d*
Intermediate SIM Hours	334	525.50	22.75	.000***	3.45[a]
Advanced SIM Hours	358	179.50			
Intermediate Aircraft Hours	336	278.65	8.77	.000***	0.71
Advanced Aircraft Hours	358	412.12			

***$p < .001$ [a] Cohen's d = very large effect size ($d \geq 1.30$)

The mean rankings of flight simulation training hours of intermediate and advanced pilot trainees were significantly different. The intermediate pilot trainees utilized simulation training significantly more often ($p < .001$) than advanced pilot trainees, and the magnitude of effect as measured by *Cohen's d* was considered very large ($d > 1.30$).

With regard to actual aircraft training time, the mean rankings between intermediate and advanced trainees were also significantly different ($p < .01$). Advanced pilot trainees utilized actual aircraft time significantly more often than intermediate pilot trainees, and the magnitude of effect (effect size) in this comparison approached large ($d = .80$). These results point to the greater use of simulation time at the intermediate level of pilot training than at the advanced level.

Due to the statistically significant finding for differences in simulation and flight training mean rankings of intermediate and advanced pilot trainees, the null hypothesis ($H_0 3$) for Research Question 3 was rejected.

Ancillary Results

Based on the significant findings of the intermediate and advanced groups favoring actual aircraft training time in Research Questions 1 and 2, additional analyses were conducted to compare the intermediate and advanced training groups' scores on the NSS to determine whether significant differences existed. Scores on the NSS were disaggregated into 3 different NSS sets (NSS ≤ 50, NSS +1 Standard Deviation (*SD*), and NSS + 2 *SD*) then compared based on trainee group. Table 6 depicts the results of the comparisons of the two training groups using the *Chi-Square* test for the NSS category of NSS ≤ 50.

Table 6

Chi-Square Comparison of NSS Score 50 ≤ *by Pilot Training Group*

Group	NSS 50 ≤ (*n*)	NSS 50 ≤ (%)
Intermediate	140	42.4
Advanced*	182	51.0

*x^2 (1) = 5.94 p = .03

The results revealed that the two training groups in this comparison were significantly different ($p < .05$) in favor of the advanced group. In other words, advanced pilot trainees were significantly more likely than intermediate pilot trainees to score at the NSS mean score of 50 or below.

Tables 7 and 8 depict the results of comparisons of NSS scores of the intermediate and advanced training groups.

Table 7

Chi-Square Comparison of NSS Score +1 SD≤) by Pilot Training Group

Group	NSS 60 (+1 *SD*) ≤ (*n*)	NSS 60 (+1 *SD*) (%)
Intermediate	49	14.8
Advanced	55	15.4

x^2 (1) = 0.04 p = .84

Table 8

Chi-Square Comparison of NSS Score +2 SD≤ by Pilot Training Group

Grouping	NSS 70 (+1 *SD*) ≤ (*n*)	NSS 70 (+1 *SD*) (%)
Intermediate	17	5.2
Advanced	10	2.8

x^2 (1) = 2.51 p = .11

Tables 7 and 8 reveal that there were no significant differences between intermediate and advanced pilot trainees on the NSS + 1 *SD* or on the NSS + 2 *SD* scores. No statistical differences were observed between intermediate and advanced flight students who scored above the mean score of 50 on the NSS.

Summary

Actual aircraft training time was found to be a significant predictor of NSS scores for both intermediate and advanced pilot trainees. In addition, there were significant differences between mean flight simulation training times and mean actual aircraft times of intermediate and advanced pilot trainees in favor of actual aircraft training time.

Ancillary analyses were conducted to further investigate the significant findings related to NSS scores. The intermediate and advanced training groups' NSS scores were disaggregated into three groups and compared using *Chi Square*; scores at the mean of 50, 60 (NSS+1 *SD*), and 70 (NSS + 2 *SD*) were examined. Significant differences between intermediate and advanced pilot trainees were observed for NSS $\leq$ 50, but no statistical significance was found for NSS + 1 *SD* and NSS + 2 *SD*. A discussion of the results of this study follows in Chapter V.

V. DISCUSSION

The purpose of the study was to investigate the relationships between flight simulation training, actual aircraft training, and intermediate and advanced jet pilot competence as measured by the Naval Standard Score (NSS). The research design of the study was non-experimental, correlational, causal-comparative with an emphasis upon the establishment of mathematical and predictive relationships using archival data**.** This chapter includes an overview of the study and its results, implications of the results, recommendations for future research, and significance of the study.

Three research questions along with three null hypotheses were formally posed to address the stated research problem of the study. After approval by Southeastern University's IRB and the Navy's IRB, archival data were obtained from the CNATRA TIMS database for years 2015 to 2017. The sample size consisted of 334 intermediate flight students and 358 advanced flight students from Naval Training Wings ONE and TWO. Research Questions one and two investigated whether flight simulation training time or actual aircraft training time was the best predictor of NSS scores of intermediate and of advanced pilot trainees. The independent variables were actual aircraft hours and flight simulation hours. The dependent variable was the NSS score.

NSS scores are calculated separately for each flight student who completes each phase of training (primary, intermediate, and advanced). This study investigated only flight students in the intermediate and advanced phases of training. The NSS is a cumulative score based on training events' scores. The score is centered on 50 with a standard deviation normalized to 10.

To address the first and second Research Questions, multiple linear regression analyses were used to determine whether flight simulation training time and actual aircraft training times were significant predictor(s) of NSS scores of intermediate and of advanced pilot trainees. The independent variables in each prediction model were flight simulation training time in hours and actual aircraft training time in hours. The dependent variables in each prediction model were NSS scores of intermediate and of advanced pilot trainees.

Research Question three focused upon differences between intermediate and advanced pilot trainees' flight simulation times and actual aircraft times. Because the data arrays were non-normal, the Mann-Whitney U test was used in Research Question three to evaluate the differences between mean ranks of training type (flight simulation time versus actual aircraft time) and training group (intermediate and advanced).

Ancillary analyses were also conducted to compare the intermediate and advanced training groups' mean scores on the NSS to determine whether significant differences existed. Scores on the NSS were disaggregated into three different NSS sets (NSS ≤ 50, NSS +1 Standard Deviation (*SD*), and NSS + 2 *SD*) and then compared using the Chi-Square test for both intermediate and advanced pilot training groups.

Overview of the Results

The mean flight simulation time for the intermediate training group was 84.72 hours; the mean simulation time for the advanced training group was 24.92 hours. The mean actual aircraft time for the intermediate training group was 81.34 hours; the mean actual aircraft time for the advanced training group was 85.64 hours. The demographic data revealed that intermediate pilot trainees utilized simulation training more than advanced pilots, while both intermediate and advanced pilot trainees flew approximately equal numbers of hours in actual aircraft.

Research Question one was predictive in nature; the results indicated that actual aircraft training time was a strong, significant predictor ($p < .001$) of NSS scores for intermediate pilot trainees. Additionally, the difference in correlations between flight simulation training time and actual aircraft training time with regard to NSS scores of intermediate pilots was statistically significant ($z = -4.10$; $p < .001$). The magnitude of effect favoring actual aircraft training time was considered medium (Cohen's $q = 0.32$) (Cohen, 1988).

Research Question two was also predictive in nature; the results indicated that actual aircraft training time was a significant predictor ($p < .05$) of NSS scores for advanced pilot trainees. The difference in correlations between flight simulation training time and actual aircraft training time with regard to NSS was not statistically significant ($z = 1.48$; $p = .07$). The magnitude of effect favoring actual aircraft training was considered small (Cohen's $q = 0.11$) (Cohen, 1988). Though actual aircraft training time was found to be a significant predictor of NSS scores for advanced pilot trainees, the difference in correlations (flight simulation versus actual aircraft) was not statistically significant. The results reveal that actual aircraft training time significantly predicts NSS scores for advanced pilot trainees; however, the correlational relationships between each training type (flight simulation and actual aircraft) and the NSS were not statistically significant.

Research Question three focused on whether any significant differences existed between the mean ranks of NSS scores based on training type (simulation and flight) and training groups (intermediate and advanced). The analyses revealed that intermediate pilot trainees utilized flight simulation training significantly more often ($p < .001$) than advanced pilot trainees. The magnitude of effect was considered very large (*d* > 1.30). Comparisons of actual aircraft training time revealed that the mean rankings between training groups were also significantly different (*p* < .01). The results suggest that advanced pilot trainees utilized actual aircraft time significantly more often than intermediate pilot trainees. The magnitude of effect approached large (*d* = .80).

Ancillary analyses were conducted on the NSS cutoff scores in three sub-groups: NSS ≤ 50, NSS +1 Standard Deviation (*SD*), and NSS + 2 *SD*. The results of the comparisons indicated that the advanced pilot trainees were significantly more likely (*p* < .05) than intermediate pilot trainees to score at the NSS mean of 50 or below. Furthermore, no statistical significance was found between intermediate and advanced pilot trainees for NSS + 1 *SD* or for NSS + 2 *SD* groups.

Implications of the Results

The results for Research Questions one, two, and three were consistent with literature based on Bloom's (1956) comprehensive model for higher-order thinking and the pilot training tasks involved in the US Navy's T-45C intermediate and advanced training phases. Intermediate military pilot training consists of basic instruments, air navigation, cockpit familiarization, basic formation flying, and runway carrier take-off and landing practice (Naval Air Training Command, 2014). These tasks involve the knowledge, comprehension, application, and analysis levels of Bloom's adapted model for pilots theorized by the researcher. Advanced military pilot training consists of operational navigation, tactical maneuvering, weapons delivery, advanced formations, low level flying, and aircraft carrier qualification (Naval Air Training Command, 2014). Realistic scenario-based training is demanded in the advanced training phase compared to the intermediate training phase. The advanced phase of US Navy flight training heavily involves the *analysis*, *synthesis,* and *evaluation* levels of the Bloom model.

Since actual aircraft training time was a positive significant predictor of NSS scores for both intermediate and advanced training phases, one could logically conclude that T-45C actual aircraft training is more effective than T-45C flight simulator training. However, the magnitude of effect favoring actual aircraft training for the intermediate training phase was considered medium (Cohen's $q = 0.32$) and, the magnitude of effect favoring actual aircraft training in the advanced training phase was considered small (Cohen's $q = 0.11$). These results suggest that simulation training time may be more effective and efficient in the intermediate phases of pilot learning than in the advanced phases.

The results of Research Question three in the current study revealed that advanced pilot trainees spent significantly more training time in actual aircraft than intermediate pilots. The magnitude of effect of the comparison was large ($d = .80$), which was not surprising since the advanced training phase includes more difficult tasks than the intermediate training phase. These results align with those of Betts et al. (2010) who found that pilots ranked Close Air Support (CAS) or strike as the most critical mission training tasks for deployed military pilots. Close air support and strike tasks used in combat and other mission-critical scenarios involve a great deal of rehearsal to achieve optimal functioning, and the margin for error is quite small.

Ancillary analyses of the current study disaggregated NSS scores into 3 different sets (NSS $\leq$ 50, NSS +1 Standard Deviation (*SD*), and NSS + 2 *SD*) and then compared the NSS scores based on intermediate or advanced trainee group. A NSS of 50 or below is considered below average and not passing. The results of the ancillary analyses indicated that the difference between intermediate and advanced pilots at the NSS cutoff of 50 or less was significant ($p = .03$). Advanced pilot trainees were significantly more likely than intermediate pilot trainees to score at the NSS mean of 50 or below. Again, the results favoring advanced pilot trainees is not surprising since the advanced phase of pilot training is more difficult than intermediate training. As a result, more flight students in the advanced phase earned a NSS of 50 or below; a score below 50 is the Navy's cutoff score for not passing each level of flight training. When the NSS scores were one or two standard deviations above the mean, there were no significant differences between intermediate and advanced pilot trainees on the NSS. However, the numbers and percentages of pilot trainees who scored one or two standard deviations above the mean on the NSS were small, indicating the inherent difficulty of the tasks and pilot execution required to earn a score above the mean.

Whatever the optimal ratios of flight simulation time to actual aircraft time, the review of literature points to the critical need for simulators that are designed for maximum fidelity in order to contribute to pilot competence (Povenmire & Roscoe, 1972). The RAND study conducted by Schank et al. (2002) found that F/A-18 simulators were not used or widely accepted by pilots due to low fidelity; however, P-3C simulators were used more often because the P-3C flight simulators more closely replicated the actual P-3C flight environment.

Flight simulation has definite value in flight training. According to a study conducted by Koglbauer et al. (2016), flight students' aircraft performance scores were significantly higher on a post-test ($p < .01$) after using a flight simulator. The authors concluded that a combination of actual flight training and flight simulator training would have positive effects on pilots' competence.

Conducting simulator flights prior to actual flights reinforces positive transfer of skills. Roof (1996) found that significant cost savings, a critical concern for military budgets, could be realized when military training events were flown in a flight simulator more than 20 times. These findings were also uncovered at a time when flight simulations were designed for a PC and were not especially sophisticated compared to today's flight simulators. Roof also concluded in his 1996 study that certain flight tasks could be effectively moved to simulators without compromising pilot training; however, he was quick to point out that critical "must-fly" tasks he identified and mapped should be carried out in actual aircraft training.

Schank et al. (2002) also suggests that independent of simulation accuracy or fidelity, some training events such as tasks requiring essential physiological cues are best flown in an actual aircraft. The T&R matrices developed by the military, combined with the "must-fly" criteria developed by Roof (1996), will provide important design information to guide the development of high-fidelity flight simulations.

Recommendations for Future Research

The current study focused on flight simulator and actual aircraft training times for one type of aircraft, the T-45C. The study could be replicated using other military aircraft types and simulators, then repeated for commercial aviation. Additionally, the study did not assess individual training event scores, but evaluated only overall pilot NSS scores of intermediate and advanced military pilot trainees. Additional studies could focus on comparing pilot trainees' scores on individual training tasks to determine whether a specific training task score could be predicted based on flight simulation time or actual aircraft training time. This type of study might advance the research evidence needed to determine the optimal ratios of simulation time versus aircraft time for military pilots without compromising pilot readiness and competence.

The researcher was privileged to conduct observations of pilot training events and conversations with flight instructors at the US Navy Training Air Wing TWO and the researcher's observations at Training Air Wing TWO. The results of the researcher's observations and conversations can be summarized in the pilots' recommendations for future research: (a) improvement of flight simulation fidelity in the intermediate flight training phase; the T-45C flight simulator plays an important role in familiarizing the intermediate pilot with the cockpit hardware and aircraft systems' functionality, and designers need to continue to improve and assess the quality of the simulations; (b) improvement of T-45C flight simulator visuals to prepare the advanced pilot trainee to be effective at formation flying, tactics, and weapons delivery; (c) investigations of the possibility of including Virtual Reality (VR) and Augmented Reality (AR) in the advanced phases of flight simulator training to make visuals outside the cockpit more realistic; (d) research on new methods to enhance de-briefs after simulator or flight training events; and (e) research on the possibility of 3-dimensional capabilities to help de-brief specific training scenarios. Research related to these specific areas of need will advance military pilot training in valuable ways.

Based on the review of literature, flight training programs should encourage greater simulator participation by pilots at all stages of their training and development and make better use of simulators as a complement to actual aircraft training, but not as a substitute for actual aircraft training. Roof (1996) concluded that substantial cost savings could be realized by moving certain tasks, excluding "must-fly" tasks, from live aircraft training to a simulator with no degradation in training or safety. These conclusions require further study and analysis.

Very few experimental studies have been conducted to compare the effectiveness of flight simulation training versus actual aircraft training. However, measuring performance and learning during flight training is essential. Carretta and Dunlap (1998) conducted a study that found that pilots with 40 to 60 simulator sessions demonstrated significantly better landing skills than pilots with 20 simulator sessions. The Carretta and Dunlap (1998) study should be replicated by the military since flight simulators have advanced in technology.

Future research should examine methodologies and evaluation tools that could be used to conduct flight simulation versus actual aircraft military flight training experiments. For example, a thorough analysis of the Navy's Training and Readiness Matrices for aircraft types should be carefully reviewed. Using the analyses and experimental results, tasks from the actual aircraft could be mapped and built into the design of a flight simulator. Mapping essential aircraft training tasks to simulator devices is an important step in ensuring that operational readiness is not compromised.

Simulation fidelity remains a concern; based on Bloom's cognitive model, higher-order thinking is critical when simulating flight as well as when flying an aircraft. Even with advances in electronic technologies such as high-definition and faster computer processing, simulation models should incorporate higher fidelity to help support more training tasks. VR and AR are advanced technologies that have the potential to be highly effective in flight simulation training. For example, AR may bring value added both in cost savings and fidelity to current simulators without the extensive need for physical or software flight simulator changes. This type of research will undoubtedly prove highly useful to both military and commercial pilot training programs.

More research should be conducted on pilots' views on flight simulation and the concern that many pilots do not take simulation practice as seriously as actual aircraft flight (Schank et al., 2002). Flight instructors should be well trained in the use of the simulation and could incorporate a method that holds pilots accountable for simulation training. Perhaps the simulations can be timed to more realistically emulate rapid and error-free decision-making required of every pilot. Flight instructors could require pilots to repeat tasks in the simulator until the task is executed perfectly. The Navy may wish to consider increasing the required number of hours spent in simulators as their design elements improve. As the need arises to move more aircraft training tasks into the simulator, the accountability of both pilots and instructors is vitally important not only for learning, but also for flight safety. Research designed to address these important recommendations will provide more evidence regarding the critical design and pilot usage elements that make simulations effective.

Significance of the Study

The Department of Defense, Federal Aviation Administration, and the International Civil Aviation Authority all seek ways to evaluate the capabilities of flight simulations and to discover methods they can be used to improve pilot training and reduce costs. This study provides further evidence on flight simulation training time versus actual aircraft training time and its impact on overall measures of pilot effectiveness to aviation authorities, flight instructors, the military aviation community, the commercial aviation community, and academia. The results of the study may lead to a better understanding of the role that flight simulation training hours and aircraft training hours play in preparing student pilots. The study may also help the military, government, and civilian authorities establish criteria for the relative amount of flight simulation time and aircraft time to incorporate into pilot training programs. Finally, the information provided by the study may ultimately help pilot trainers and simulation designers to evaluate flight training programs, increase the numbers of successful flight student completers, and accomplish the military's mission of bringing pilots to the highest levels of military preparation.

References

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

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete Edition). New York, NY: Longman.

Beaubien, J., Stacy, W., Wiggins, S., & Lucia, L. (2016). *How can we measure learning? Let's count the ways!* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Bell, B., Tannenbaum, S., Ford, J., Noe, R., & Kraiger, K. (2017). *100 years of training and development research: What we know and where we should go.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Betts, L. R., McCauley, E. M., & Walwanis, M. (2010). *Assessing the need for a deployable training system for the F/A-18 hornet and super hornet.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, NY: David McKay.

Blow, C. (2012). *Flight school in the virtual environment: Capabilities and risks of executing a simulations-based flight training program* [Monograph]. Fort Leavenworth, KS: School of Advanced Military Studies.

Boeing. (2018). *T-45 goshawk trainer*. Retrieved from http://www.boeing.com/history/products/t-45-goshawk-trainer.page

Cannon-Bowers, J., Salas, E., Tannenbaum, S., & Mathieu, J. (1995). Toward theoretically based principles of training effectiveness: A model and initial empirical investigation. *Military Psychology, 7*(3), 141-164.

Carretta, T., & Dunlap, R. (1998). *Transfer of training effectiveness in flight simulation* (Research Report No. AFRL-HE-AZ-TR-1998-0078). Dayton, OH: US Air Force Research Laboratory.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cowan, N. (2010). The magical mystery of four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19,* 51-57.

Department of Defense. (2016). *Operation and maintenance overview: Fiscal year 2017 budget estimates.* Retrieved from http://comptroller.defense.gov/Portals/45/Documents defbudget/fy2017/fy2017_OM_Overview.pdf

Department of the Navy. (2010). *Simulated training strategy*. Washington, DC: Department of the Navy.

Department of the Navy. (2014). *Chief of naval operations simulator development and training strategy*. Washington, DC: Department of the Navy.

Federal Aviation Administration. (2004). *Airplane flying handbook*. Washington, DC: US Department of Transportation.

Federal Aviation Administration. (2017). *Aviation instructor's handbook*. New York, NY: Skyhorse Publishing.

Federal Aviation Administration. (2018). *Federal aviation regulations and aeronautical information manual (FAR/AIM)*. Newcastle, WA: Aviation Supplies and Academics.

Faubert, J., & Sidebottom, L. (2012). Perceptual-cognitive training of athletes. *Journal of Clinical Sports Psychology, 6*(1), 85-102.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Thousand Oaks, CA: SAGE Publications.

Fiske, D. W. (1947). Validation of naval aviation cadet selection tests against training criteria. *Journal of Applied Psychology, 31*, 601-614. doi: 10.1037/h0054274

Foster, C., Melon, E., & Philips, H., IV. (2007). Undergraduate military flight officer training: Increasing training effectiveness through job task and training media analysis. *Performance Improvement, 46*(3), 36-41.

Geissler, L. R. (1918). A plan for the technical training of consulting psychologists. *Journal of Applied Psychology*, *2*(1), 77-83. doi: 10.1037/h0072539

Hays, R. (2006). *The science of learning: A systems theory perspective.* Boca Raton, FL: Brown Walker Press.

Hirano, M., Kubota, S., Tanabe, S., Koizume, Y., & Funase, K. (2015). Interactions among learning stage, retention, and primary motor cortex excitability in motor skill learning. *Brain Stimulation, 8*(6), 1195-1204. doi: 10.1016/j.brs.2015.07.025

Hoke, J., Reuter, C., Romeas, T., Montariol, M., Schnell, T., & Faubert, J. (2017). Perceptual-cognitive and physiological assessment of training effectiveness. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Jacobs, J., Prince, C., Hays, R., & Salas, E. (1990). A meta-analysis of the flight simulator training research [Technical Report No. 89-006]. Retrieved from www.dtic.mil/dtic/tr/fulltext/u2/a228733.pdf

Kellogg, W. N. (1946). The learning curve for flying an airplane. *Journal of Applied Psychology, 30*(5), 435-441. doi: 10.1037/h0060547

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *American Society for Training and Development, 13*(11), 3-9.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook* (2nd ed.). New York, NY: McGraw-Hill.

Koglbauer, I. (2016). Simulator training improves pilots' procedural memory and generalization of behavior in critical flight situations. *Journal of Cognition, Brain, and Behavior, 20* (5), 357-366.

Koglbauer, I., Riesel, M., & Braunstingl, R. (2016). Positive effects of combined aircraft and simulator training on the acquisition of visual flight skills. *Journal of Cognition, Brain, and Behavior, 20*(5), 309-318.

Molesworth, B. R. C., Bennett, L., & Kehoe, J. (2010). Promoting learning, memory, and transfer in a time-constrained, high hazard environment. *Accident Analysis and Prevention, 43*(3), 932-938. doi: 10.1016/j.aap.2010.11.016

Naval Air Training Command. (2014). *Naval flight student training administration manual:* CNATRAINST 1542.167A. Corpus Christi, TX: Chief of Naval Air Training.

Palmer, J. (2001). *Fifty modern thinkers on education: From Piaget to the present.* New York, NY: Routledge.

Pfeiffer, M. G., Horey, J. D., & Butrimas, S. K. (1991). Transfer of simulated instrument training to instrument and contact flight. *International Journal of Aviation Psychology, 1*(3), 219-229.

Povenmire, K. H., & Roscoe, N. S. (1972). *The incremental transfer effectiveness of a ground-based general aviation trainer* (Research Report No. ARL-72-9/AFOSR-72-4). Savoy, IL: Aviation Research Laboratory.

Rogers, R., Boquet, A., Howell, C., & DeJohn, C. (2007). *Preliminary results of an experiment to evaluate transfer of low-cost, simulator-based airplane upset-recovery training.* Washington, DC: Federal Aviation Administration.

Roof, R. (1996). *Naval aviation's use of simulators in the operational training environment: A cost analysis perspective* [Master's thesis]. Retrieved from Naval Post Graduate School archive.

Schank, J. F., Thie, H. J., Graff, C. M., II., Beel, J., & Sollinger, J. (2002). *Finding the right balance: Simulator and live training for navy units.* Santa Monica, CA: RAND.

Smith-Jentsch, K. A., Jentsch, F. G., Payne, S. C., & Salas, E. (1996). Can pre-training experiences explain individual differences in learning? *Journal of Applied Psychology, 81*(1)*,* 110-116.

Stevens-Adams, S., Basilico, J., Abbott, R., Gieseler, C., & Forsythe, C. (2010). Performance assessment to enhance training effectiveness. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Sousa, D. (2017). *How the brain learns* (5th ed.). Thousand Oaks, CA: Corwin.

Swezey, R. W., Owens, J. M., Bergondy, M. L., & Salas, E. (1998). Task and training requirements analysis methodology (TTRAM): An analytic methodology for identifying potential training uses of simulator networks in teamwork-intensive task environments. *Ergonomics*, *41*(11), 1678-1697. doi: 10.1080/0014013986135

US Congress. (2017). Admiral Moran state of the military. *Proceedings of the US House Armed Services Committee*, Washington, DC.