NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

TACHISTOSCOPE ON A VIRTUAL REALITY PLATFORM TO IMPROVE MEMORIZATION AND INCREASE RAPID RECOGNITION

by

Brian T. Pugh

June 2022

Thesis Advisor:	Quinn Kennedy
Second Reader:	Dorion Liston (NPS)

Research for this thesis was performed at the MOVES Institute.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC, 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2022	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE TACHISTOSCOPE ON A VIRTUAL REALITY PLATFORM TO IMPROVE MEMORIZATION AND INCREASE RAPID RECOGNITION			**5. FUNDING NUMBERS**
6. AUTHOR(S) Brian T. Pugh			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) ONR, Arlington, VA 22201			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

This work investigates whether a tachistoscope on a virtual reality (VR) platform can increase one's ability to memorize and rapidly recognize objects. These abilities are relevant to an array of military requirements. Current procedures mostly utilize flash cards and PowerPoint slides. A tachistoscope (ta-kiss-stow-scope) is an image-flashing device with precise control of the image presentation time. Since the early 1900s they were used to assist with memorization and recognition. One famous example is work done by Renshaw in the 1940s to improve pilots' ability to recognize tanks, aircraft, and ships (Renshaw, 1945). Our study utilized this technique on modern-day VR and computer platforms. It simplified the use of a tachistoscope and will enable units to customize training packages. This study trained individuals to recognize 40 aircraft over eight training sessions. Training session one began with ten aircraft, and five aircraft were added in each subsequent session. Questions captured three variables: correct/incorrect answer, reaction time, and confidence. Participants were in one of three groups: tachistoscope on a VR platform, tachistoscope on laptop, or computer-based flashcard (control). Results indicate a significant increase in memorization from pretest to posttest for all groups. Furthermore, there was a nonsignificant improvement in reaction time from pretest to posttest across all groups.

14. SUBJECT TERMS memorization, rapid recognition, tachistoscope, human performance, virtual reality, threat vehicle identification, TVI, naturalistic decision making, recognition-primed decision making			**15. NUMBER OF PAGES** 105
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

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

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

TACHISTOSCOPE ON A VIRTUAL REALITY PLATFORM TO IMPROVE MEMORIZATION AND INCREASE RAPID RECOGNITION

Brian T. Pugh
Major, United States Marine Corps
BS, University of Florida, 2005

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MODELING, VIRTUAL ENVIRONMENTS, AND SIMULATION

from the

NAVAL POSTGRADUATE SCHOOL
June 2022

Approved by: Quinn Kennedy
Advisor

Dorion Liston
Second Reader

Gurminder Singh
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This work investigates whether a tachistoscope on a virtual reality (VR) platform can increase one's ability to memorize and rapidly recognize objects. These abilities are relevant to an array of military requirements. Current procedures mostly utilize flash cards and PowerPoint slides. A tachistoscope (ta-kiss-stow-scope) is an image-flashing device with precise control of the image presentation time. Since the early 1900s they were used to assist with memorization and recognition. One famous example is work done by Renshaw in the 1940s to improve pilots' ability to recognize tanks, aircraft, and ships (Renshaw, 1945). Our study utilized this technique on modern-day VR and computer platforms. It simplified the use of a tachistoscope and will enable units to customize training packages. This study trained individuals to recognize 40 aircraft over eight training sessions. Training session one began with ten aircraft, and five aircraft were added in each subsequent session. Questions captured three variables: correct/incorrect answer, reaction time, and confidence. Participants were in one of three groups: tachistoscope on a VR platform, tachistoscope on laptop, or computer-based flashcard (control). Results indicate a significant increase in memorization from pretest to posttest for all groups. Furthermore, there was a nonsignificant improvement in reaction time from pretest to posttest across all groups.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ANOVA	Analysis of Variance
AWS	Amazon Web Services
COA	Course of Action
HummRRO	Human Resource Research Office
ISI	Interstimulus Time Interval
NDM	Naturalistic Decision Making
ROC-V	Recognition of Combat Vehicles
RSVP	Rapid Serial Visual Presentation
RTS	Renshaw Training System
SME	Subliminal Mere Exposure
TVI	Threat Vehicle Identification
VR	Virtual Reality
WEFT	Wings, Engines, Fuselage, and Tail

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

Writing a thesis is a team effort, it cannot be done alone. First, I would like to thank God for giving me this opportunity and putting people in place to help me along the way. Without Him this would not be possible. Second, I would like to thank my wife, Sofia, and children, Theodore, Elliott, and Amelia. Prior to my arrival at NPS, I did not understand how difficult this program would be. Sofia, you served as mother, wife, teacher, secretary, manager, cook, and waitress. Without you I would not have eaten or slept, thank you. To my children, thank you for keeping things exciting and being part of my Zoom classes. You sacrificed a lot of Daddy time to make this happen. We will make it up in the camper.

Dr. Quinn Kennedy, from our first conversation you were extremely supportive and encouraging. You kept me on track and would not let me waver. I appreciate all your experience and mentorship. While your standards are very high, you always set me up for success. I could not have asked for a better advisor.

Dr. Dorion Liston, without you I would still be writing scripts. Thank you for your one-on-one mentorship and assistance. Your work is amazing, I am excited to see where it takes you in the future.

Chris Fitzpatrick, you are the man. When there is a problem or challenge you have been there for us. You provided the hardware to make this study happen. Thank you for relieving stress and supporting when needed.

Future Tech Team, without you I would still be building the Retention Software Program. You do amazing work and NPS is lucky to have you.

ONR, thank you for believing in my idea and funding this project.

To my cohort, it has been a ride. It is hard to imagine how much we have learned in the past two years. It was not easy, but we made it. I appreciate all your support and look forward to working with you in the fleet.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

The purpose of this thesis is to develop a software system to run a tachistoscope on a virtual reality and computer-based platform, and to determine if that software can increase a learner's ability to memorize and rapidly recognize objects (e.g., aircraft, vehicles, people, etc.). The basis of this research is founded on two bodies of work. First, the research and training conducted by Samuel Renshaw in the 1940s that utilized a tachistoscope to help U.S. Navy and Army pilots more accurately and quickly identify aircraft and ships during World War II (Renshaw, 1945). Second, Gary Klein's recognition-primed decision-making model used in Naturalistic Decision Making (G. Klein, 2008).

1. Memorization and Rapid Recognition

The ability to rapidly recognize and memorize objects is relevant to an array of military requirements. For instance, the recognition of vehicles and aircraft is essential to sort clutter on the battlefield. The challenge of vehicle recognition is compounded in low-light and low-visibility situations. "During Operation DESERT STORM, approximately 80% of M1 Abrams Tanks and M2 Bradley Fighting vehicles were destroyed by friendly fire" (Rowan, 2015). Further compounding the challenge of vehicle recognition is the fact infrared and thermal optics were used in 90% of combat engagements (Rowan, 2015). The Army evaluates a soldier's ability to conduct threat vehicle identification (TVI) using its gunnery skills test in accordance with Army Field Manual 3-20.3. One purpose of TVI is to decrease the risk of fratricide. In the conduct of the test, soldiers are presented with an image on a PowerPoint slide for 10 seconds. The soldier then has 10 more seconds to correctly identify the vehicle or aircraft. Soldiers are required to identify 50 vehicles, 25 under daylight conditions and 25 under thermal conditions (*Tc3_20x31_1.Pdf*, n.d.). The object identification test in this study is modeled after the TVI test.

After discussion with military colleagues and personal experience, it was found many organizations use PowerPoint presentations, flash cards, and playing cards in an attempt to increase memory and rapid recognition. Memorization, recognition, and retention of objects, such as military equipment, vehicles, and personnel, is a time consuming and difficult challenge. Compounding the difficulty is individual military fields use different learning and memorization methods, but not enough effort is done to teach military members efficient methods for learning and memorization. In many settings (e.g., schoolhouse, on-the-job training, etc.), members may resort to less efficient memory methods such as staring at images for an indefinite time.

In 2013, the Army created the Recognition of Combat Vehicles (ROC-V) program to aid in this task (Rowan, 2015). ROC-V is a web-based platform that has a large database of vehicle images with descriptions of the vehicles. Managers can create custom training and testing packages to aid users in training and evaluation. Rowan (2015) showed the ROC-V platform was effective at training service members to 90% recognition, but there was a significant memory decay if the task or information was not maintained. In addition, Rowan tested participants on a relatively small data set (12 vehicles) while Participants trained for an indefinite period of time (Rowan, 2015). ROC-V was a good step in the right direction. However, there is a need for more efficient, flexible, and lasting methods to remember and rapidly recognize vehicles, equipment, and aircraft. I propose a tachistoscope on a VR platform will provide a low-cost, reliable, and flexible solution. This system would allow users to customize the training set for any type of object (e.g., vehicle, aircraft, parts, tools, people, etc.). It also provides a platform that can integrate adaptive training algorithms to provide more targeted training in less time.

2. A Possible Solution: The Tachistoscope

The word tachistoscope is derived from a combination of Greek words “takhistos” (swiftest), and “skopeo” (look at) (Milgram, 1986). In its essence, a tachistoscope is an image-flashing device with precise control on the amount of time images or stimuli are presented to a learner (Milgram, 1986). The rapid presentation of a stimulus provides two advantages in training. First, it hyper-focuses the attention of the learner because they know the stimulus will only be presented for a fraction of a second. Second, the stimulus is presented too briefly to be processed by the image forming area of the visual system, the occipital cortex. Instead, the image goes to the non-image forming portions of the visual system. This causes the learner to perceive portions of the object at a level below the threshold of conscious perception (M. Kahrhoff, personal communication, May 28, 2021). I propose this process will support the use of the recognition-primed decision model described in the Chapter II, Literature Review (G. Klein, 2008).

Tachistoscopes have been used since the mid-1850s in various forms. One of the earliest tachistoscopes was used by Wilhelm Wundt in 1897 and can be seen in Figure 1. Figure 2 provides an example of Samuel Renshaw in the lab using his tachistoscope. Users would look in a viewer and then either a shutter was rapidly opened and closed, or an image was dropped in front of their view for a brief amount of time depending on the type of tachistoscope. In following years, tachistoscopes took multiple forms to include wheels and spectacle-mounted liquid-crystal displays that use occlusion to block a learner's view (see Figure 3) (Milgram, 1986).

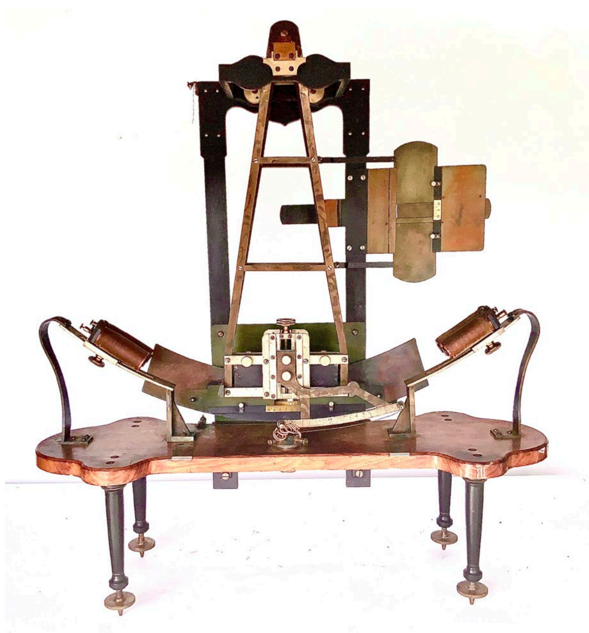


Figure 1. 1897 Wilhelm Wundt Tachistoscope. Source: *Fleaglass* (n.d.).

Figure 2. Renshaw with Tachistoscope. Source: Rightbraineducation (2019).

Figure 3. Electronic Tachistoscope. Source: Rightbraineducation (2019).

Eventually tachistoscopes started using electrical lights and more recently computer displays (Bauer, 2015). At first glance, computer displays appeared to be an ideal platform. However, until recently, monitors and screens had substandard refresh rates which caused issues with the precise time required to present a stimulus (Sperdin et al., 2013). Using a monitor introduced another potential issue not seen in the traditional tachistoscope. Instead of looking through an enclosed viewer, individuals would look at a screen and while exposed to the external confounding visual stimuli in an individual's peripheral vision. Applying the tachistoscope on a VR platform could address these issues.

Current refresh rates on VR headsets have increased to levels that can execute precise presentation speeds (72 hz) (*Oculus Device Specifications | Oculus Developers*, n.d.). In addition, the headsets block out all other visual stimuli thus decreasing confounding stimuli and increasing focus.

B. RESEARCH QUESTIONS

1. Standard Method

Participants in the Control Group used what will be called the Standard Method. The Standard Method will use the same software designed to train the experimental groups, the Retention Software Program. To simulate the use of flashcards, images will be presented on a computer screen one time for five seconds.

2. Research Question 1

Can a tachistoscope (Tscope) on a virtual reality platform improve an individual's ability to memorize an object (aircraft) as compared to a tachistoscope on a computer screen and the standard method?

HO1: There is no difference between the Control Group (standard method), Tscope on Laptop group, and the Tscope on the VR Headset Group in the average increase in the number of objects identified before and after training.

$\mu_{\text{control-d}} = \mu_{\text{VRTscope-d}} = \mu_{\text{CompTscope-d}}$.

HA1: There is a difference between the Control Group (standard method), Tscope on Laptop Group, and the Tscope on the VR Headset Group in the average increase in the number of objects identified before and after training.

$\mu_{\text{control-d}} \neq \mu_{\text{VRTscope-d}} \neq \mu_{\text{CompTscope-d}}$.

3. Research Question 2

Can a tachistoscope on a virtual reality platform improve the speed with which an individual recognizes an object (aircraft) as compared to a tachistoscope on a computer screen and the standard method?

HO2: There is no difference between the Control Group (standard method), Tscope on Laptop group, and the Tscope on the VR Headset Group in the average change in speed at which objects are correctly identified before and after training.

$\mu_{dRT1} = \mu_{dRT2} = \mu_{dRT3}$

b. HA2: There is a difference between the standard method group (control), computer Tscope group, and the Tscope on the VR platform group in the average change in speed at which objects are correctly identified before and after training.

$\mu_{dRT1} \neq \mu_{dRT2} \neq \mu_{dRT3}$

II. LITERATURE REVIEW

A. NATURALISTIC DECISION MAKING

1. Recognition-Primed Decision Model

Naturalistic Decision Making and the Recognition-Primed Decision Model were the result of a study conducted by Gary Klein and Klein Associates Inc. for the U.S. Army (1988). Klein was investigating decision making in an operational setting. The intent was to describe how people make decisions in real-world settings, specifically military decision makers. The program was conducted over three years and included seven separate projects. It included studies on command-and-control decision-making, decision making by a wildland fire incident commander, and the management of an Army tank platoon (G. Klein & Calderwood, 1996).

Klein observed that contrary to contemporary thought, people did not create a list of courses-of-action (COA) and then choose the best option from that list. Rather, people took in stimuli from the environment and compared it to prior experiences to build a single COA, as the model depicts in Figure 4. As the decision maker builds his or her situational assessment, they are scanning for relevant critical cues. Expert decision makers are familiar with which critical cues are important and can direct their attention to those relevant cues. In turn, they pay less attention to insignificant stimuli or cues that may distract novices (G. A. Klein & Calderwood, 1996). The relevant cues can be used to test expectancies within the current situation. If the relevant cues violate the current situation, the decision maker will then gather more information and reassess the situation to evaluate if the situation is familiar. If the relevant cues do not violate an expectancy, then the decision maker will confirm their evaluation of the situation and begin to cognitively simulate his or her COA (G. Klein, 2008).

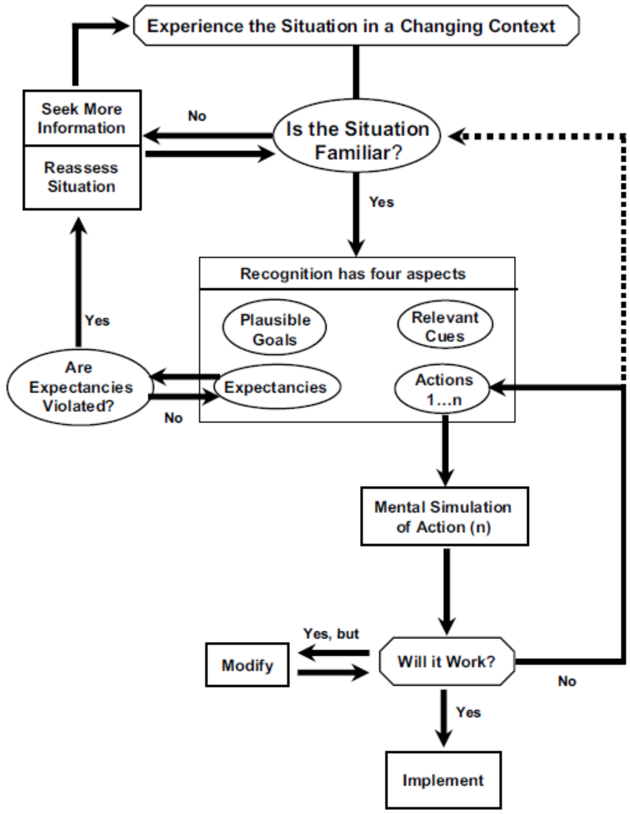


Figure 4. Recognition-Primed Decision Model. Source: G. Klein (2008).

The use of a tachistoscope on a VR platform is aligned with NDM in that it aims to encode objects into a person's implicit memory. It is conjectured the object will be intuitively recognized as a relevant cue or pattern observed in the current situation. If the object is properly encoded, the decision maker will use their tacit knowledge of the cue to determine if any expectancies are violated in their current mental model (G. Klein, 2008). If the decision maker implicitly recognizes the object, it is conjectured it can lighten their cognitive load and increase the speed at which the decision model can be executed.

1. Priming

Priming promotes the identification of a stimulus based upon prior exposure to said stimulus. There are different types of priming: visual, subliminal, supraliminal, and response priming. Visual priming is the use of a visual stimulus without the use of other stimulus types. Subliminal priming is the presentation of stimuli below a threshold perceived by the conscious mind (< ~500 ms) intended to influence a subject's implicit memory. Supraliminal priming is the presentation of stimuli that can be perceived by the conscious mind (> ~500 ms). Priming using both supraliminal and subliminal priming have been shown to have stronger and longer lasting effects than subliminal priming alone (Elgendi et al., 2018). Response priming is when a subject is exposed to the prime and stimulus in rapid succession (Elgendi et al., 2018).

Response priming demonstrates a direct influence on the implicit memory of a learner (Cave, 1997). When a person is presented with a stimulus on subsequent occasions, the previous priming influences how the brain processes that stimulus. In a study conducted by Cave (1997) 207 subjects were presented pictures in two different sessions. 130 images were presented for naming in the first session and 200 images were presented for naming in the second session. Sixty pictures were presented for recognition in the second session, 30 new and 30 old. The second session varied in delay from 6-weeks to 48-weeks depending on which of the nine groups a subject was assigned. The priming effect was measured as the difference in reaction time when naming old and new pictures in the second session. While the difference in reaction times did shrink as the delay grew, in each of the groups the old pictures had quicker reaction times when compared to the new pictures, as seen in Figure 5 (Cave, 1997).

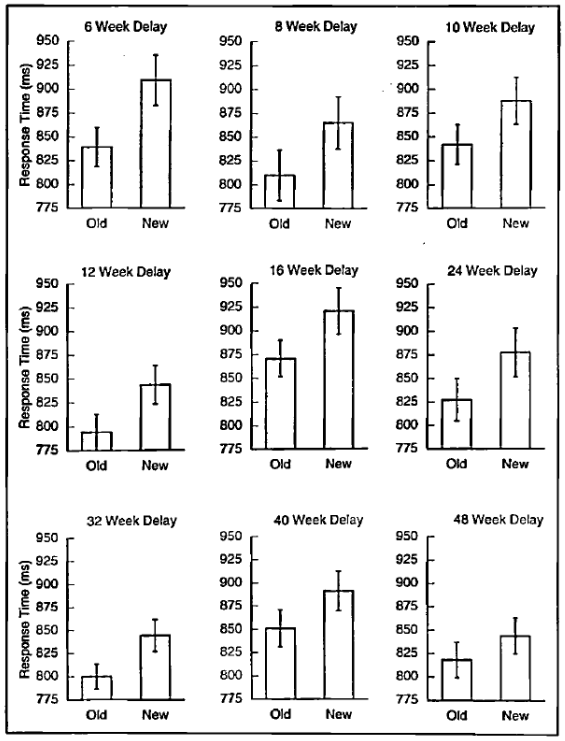


Fig. 1. Means of median times to name old and new pictures 6 to 48 weeks after the study session. Error bars represent ±1 standard error of the mean.

Figure 5. Means of Median Times to Name Old and New Picture: Very Long-Lasting Priming in Picture Naming. Source: Cave (1997).

Recognition was measured by showing the subjects 30-old and 30-new pictures in the second session. Interestingly, the decline in recognition is not monotonic and only decreases from 20% to about 15% with a 48-week delay, as seen in Figure 6. While the hit-false alarm ratio is only 20%, the response could be improved by presenting the subject with consistent primes multiple times (Cave, 1997; Elgendi et al., 2018).

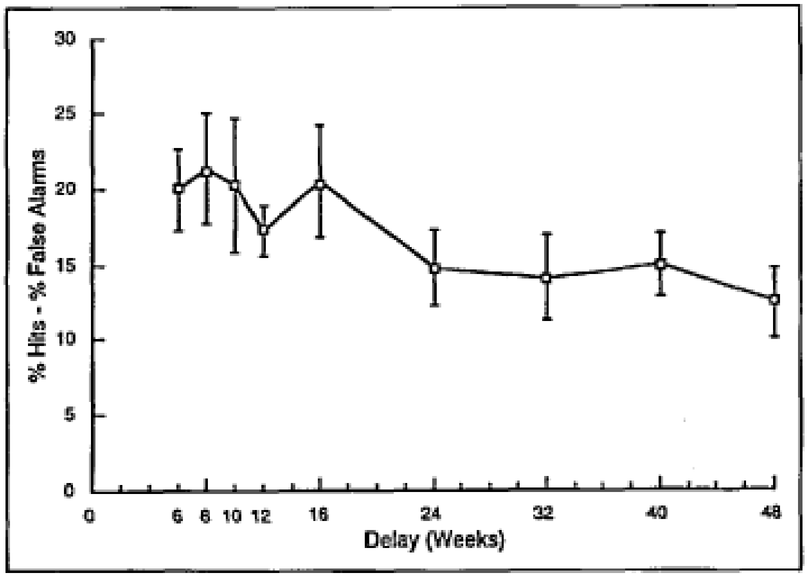


Figure 6. Recognition Performance, error bar represents +- 1 percent: Very Long-Lasting Priming in Picture Naming. Source: Cave (1997).

2. Experimental Design

Cave (1997) had two interesting methods to tests priming and recognition effects. To test priming effects, Cave measured the response time it took for subjects to respond with the name of the object in the picture and then compared the difference between new and old pictures. The second part of the test looked at recognition. Sixty additional pictures were used to evaluate the ability of individuals to recognize objects. The measurement used to evaluate was d' (the % Hits vs the % False Alarms). While the first two-weeks saw a large drop in recognition, there was little further decline in recognition over the next 42 weeks.

3. Implicit Learning and Tacit Knowledge

Implicit learning is the process of learning "dynamic statistical patterns and features" to acquire skills, knowledge, or expertise when exposed to a stimulus in a mostly unconscious manner (Patterson et al., 2010, p. 1). A tachistoscope can be used to briefly expose a learner to images with different objects (Renshaw, 1945). The exposure of images for less than 500 ms prevents the mind from consciously perceiving the individual shapes of the object (Elgendi et al., 2018). As learners are trained on the tachistoscope, they implicitly learn to aggregate the sub-units of an object into the whole, a gestalt approach. As the learner gains knowledge of the object through training, the learner requires less effort to build the mental image of the entire object (Renshaw, 1945). As this occurs, the learner moves from a declarative phase of skill acquisition, in which the learner can identify facts in a mechanical manner, to more of a procedural phase, in which the learner has task knowledge (Kim et al., 2013). If the procedural phase is reached, there is typically less forgetting and retraining can occur at a much quicker rate (Kim et al., 2013).

The end-state of the implicit learning process is tacit knowledge. Tacit knowledge is information that is stored and used in an unconscious manner (Patterson et al., 2010). The use of tacit knowledge can enhance the acquisition, retention, and transfer of more complex and explicit skills. The recognition-primed decision model is an intuitive decision process and relies heavily on implicit learning (Patterson et al., 2010). For example, when a baseball player is batting, they must make a split-second decision on whether to swing or hold. The time allowed to make such a decision does not allow the hitter to consciously run through multiple scenarios, learn about the situation, and then compare choices. The hitter must use their tacit knowledge of how a pitcher throws different pitches to immediately recognize and act. Tacit knowledge can be acquired if the hitter is implicitly and explicitly trained on the cues of particular pitches (e.g., how a pitcher holds the ball in their glove or the posture of the pitcher). That tacit knowledge can be used by the hitter to intuitively predict the type of pitch (D. Teig, personal communication, April 2021). When learned, this type of tacit knowledge can be used as tacit scaffolding to build expertise (e.g., recognizing a pitch and hitting a baseball) (Patterson et al., 2010).

It is suggested that immersive environments enhance the implicit learning process and the building of tacit knowledge (Patterson et al., 2010). Thus, it is conjectured the tachistoscope on a VR platform may perform better than a tachistoscope on a standard computer screen because it is more immersive. The VR platform can draw the user into the environment, and block much of the ancillary visual stimulation surrounding a standard computer screen.

B. RENSHAW AND THE RENSHAW TRAINING SYSTEM FOR AIRCRAFT AND SHIP RECOGNITION

1. Renshaw

Samuel Renshaw was a psychologist in the 1930s and 1940s at the Ohio State University (OSU). He primarily conducted studies with a tachistoscope to improve number memorization, reading speed, and comprehension (Renshaw, 1945). During World War II, the military was losing personnel due to poor recognition of friendly and enemy tanks, aircraft, ships, and vehicles. In January 1942, the Bureau of Aeronautics of the Navy Department submitted a proposal to use Renshaw's tachistoscope training methods for aircraft and ship recognition (Renshaw, 1945). The training system developed by Renshaw and his team was eventually adopted as the Renshaw System or Flash System of Instant Recognition.

Results from the study showed it was an effective method to train pilots on recognition of aircraft, ships, tanks, and vehicles. After 15 sessions, 30.3% of cadets identified 90% of the planes and ships correctly at 1/75 second exposure, (N = 323). After 50 sessions 74.7% of cadets identified 90% of the planes and ships correctly at 1/75 second exposure, (N = 272) (Renshaw, 1945). Results of this study are presented in Figure 7.

TABLE 6
GROUP TACHISTOSCOPIC TRAINING—NAVY PREFLIGHT CADETS

Percent of planes* and ships correctly identified	N = 323 Cumulative per cent cadets (Session 15)	N = 272 Cumulative per cent cadets (Session 50)
100	2.5	12.5
90	30.3	74.7
80	66.6	94.6
70	88.9	99.3
62.5	——	100.0
60	96.6	——
50	98.8	——
22.5	100.0	——

*Forty slides were shown at 1/25 second on Session 15; 40 at 1/75 second on Session 50. The pretest averages of the two battalions before training showed no significant differences. Many of the views of planes used in the tests were not seen previously in the training sessions.

Figure 7. Results from Renshaw Flight Cadets Training Program After 15 and 50 Sessions. Source: Renshaw (1945).

Renshaw continued to test flashing images at different speeds and found after the presentation time reached one-second, it had a negative impact on recognition (Renshaw, 1945).

To demonstrate the recognition skill was transferable to real aircraft, Renshaw tested 400 of the trained flight cadets on real planes. Ninety percent of the cadets performed the test as well on real aircraft as they did with the aircraft pictures. However, there were issues in differentiation between the size of smaller fighters and larger bombers (Renshaw, 1945). Renshaw (1945) stated he accounted for this issue in later training, but the methods used could not be identified in the literature. The issue of size differentiation should be studied in follow-on studies but is outside the scope of this study.

In his training, Renshaw was using subliminal mere exposures (SME) to train individuals using visual priming techniques. The use of SME in subliminal visual priming has been shown to be effective when a stimulus is displayed at a rapid rate (Elgendi et al., 2018). Studies in rapid serial visual presentation (RSVP) have demonstrated viewers can identify objects when they are presented in as short as 160 ms (Potter & Fox, 2009) and gain general comprehension of categorical information (e.g., cat, dog, animal, car, etc.) down to 13 ms when the image is not masked (Potter et al., 2014).

2. Renshaw Training System for Aircraft and Ship Recognition

The Renshaw Training System for Aircraft and Ship Recognition (RTS) is a visual system designed for users to look at the whole of an object, their "total forms" (Gibson, 1947, p. 128). Initial courses typically consisted of 30 one-hour sessions over a nine-week period and included about 40 different aircraft (Gibson, 1947, p. 116). However, these parameters were not concrete and fluctuated to meet training requirements. There is literature that references up to 50 sessions with 80 aircraft types and ships ("Quick, Total, Accurate Recognition," 1943; Renshaw, 1945). Before slides of aircraft were presented the aircraft characteristics were discussed and slide reviews were done for aircraft previously presented. The trainers would present 10 to 20 planes and then had cadets identify the aircraft in writing. After each exposure, the students were allowed to look at the aircraft to make corrections (Gibson, 1947). The RTS continued to develop over time. In addition, Renshaw conjectures they have not found the capacity of memory, this skill is mostly limited by the amount of time required to train (Renshaw, 1945).

The system was found to be to so effective by the U.S. Navy, they set up the Renshaw Recognition School at OSU. By the end of World War II Renshaw had trained over 4,000 pilots from across all the branches of service (*Science: Fast Looks*, 1945). In 1943, the Army Air Force officially adopted the RTS under the name "Flash System of Instant Recognition" (Gibson, 1947, page 115).

The training outlined in the Renshaw Recognition System is more robust than the capacity of this study. Thus, it is conjectured the results of this study may not be as strong as those found in Renshaw's studies and training.

2. Counter Arguments to Renshaw Training System

Between 1943 and 1945, the U.S. Army Air Corps conducted research to determine the best methods to train pilots in recognition. In 1946, Lt Col James Gibson, Air Corps, with the Psychological Test Film Unit, conducted different tests to study the theory of visual learning (Gibson, 1947). They tested cadets on aptitude and proficiency, did a review of current instructional methods at the time, and potential future methods. A detailed comparison was done on the outcomes of the different tests. Two of the main methods they compared were the Wings, Engine, Fuselage, and Tail (WEFT) and the Renshaw Training System for Aircraft and Ship Recognition (Gibson, 1947).

a. Wings, Engine, Fuselage, and Tail

WEFT is an aircraft identification training model that is in use today and dates back to the WWII period (Vicory, 1968). As implied by the name, the WEFT system focuses on individual parts of the aircraft and their unique characteristics. The goal of WEFT was to provide a systematic approach to identifying aircraft by memorizing shapes and characteristics of parts of the aircraft, an "analysis of features" (Gibson, 1947; Vicory, 1968). This technique was useful because it allowed students to study without instructors and provided the students vocabulary to describe different parts of the airplane (Vicory, 1968). In 1941, prior to the Renshaw Training System, the WEFT system was adopted by the U.S. Navy and Army Air Corps (Vicory, 1968). What the WEFT system did offer, that the Renshaw Training System did not, was the language to speak about different parts of the aircraft (e.g., “squared wing tips” or “barrel-shaped fuselage”) (Gibson, 1947, p. 128).

b. Slow, Intermediate, and Fast Presentation Testing (1943 – 1945)

Gibson, (1947) conducted a study to examine the effects of presentation speed on memory performance. There were three experimental groups, a slow-trained group (1 sec slide presentation), an intermediate-trained group (1/10 sec slide presentation) and fast-trained group (1/50 second slide presentation). Each experimental group consisted of four cadet classes with approximately 40 cadets per class, about 160 students per group (Gibson, 1947). After 30 hours of training, each group was tested on how accurately they could identify 20 aircraft presented at each of the three presentation speeds (1 sec, 1/10 sec, and 1/50 sec). All the groups performed equivalently well at each of test presentation speeds, as seen in Figure 8 (Gibson, 1947). If there is no difference in accuracy between the groups but the cadets spent less time learning, then it could be argued the fast-trained group was trained more efficiently.

TABLE 7.1.—*Proficiency in aircraft recognition of three experimental groups as measured by slide examinations at different exposure speeds*

Group	*N*	*Test at 1 sec.*		*Test at 1/10 sec.*		*Test at 1/50 sec.*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Slow-trained	173	16.7	2.2	14.3	3.0	13.0	2.9
Intermediate-trained ..	167	16.7	2.3	14.3	2.8	13.6	2.9
Fast-trained	177	17.0	2.4	14.0	2.8	14.1	3.3

Figure 8. Test Results for Cadets Trained and Tested at Slow (one-sec), Intermediate (1/10-sec), and Fast (1/50-sec). Source: Gibson (1947).

c. Counter-to-the-Counter Argument

Vicory (1967) argues the Renshaw method was “inappropriate” and some of his claims were “unsupported.” In that the Renshaw method was “inappropriate,” Vicory states there is no requirement for an observer to make such a rapid recognition; that a pilot would have sufficient time to make an identification before the enemy aircraft can get within firing range (Vicory, 1968). While this statement may have been true for pilots at the time, in the 21st century, there are many instances when an observer or gunner would need to make a rapid decision in a shoot/don't shoot situation (e.g., room clearing operations, urban operations, etc.). Regarding Vicory's assertions of Renshaw's claim of "identification skill improves more with shorter than with longer exposures," Vicory could potentially be correct once the presentations become too short, less than 100 ms. More research will need to be done to determine the effective threshold. However, this thesis study will not present images below a 250 ms threshold, which is above the intermediate exposure time (100 ms) that was shown to have to no significant difference when compared to the slow exposure time (1 sec) (Vicory, 1968). Furthermore, this study is not in conflict with a study Renshaw conducted in which he found a degradation in recognition when the image was presented for longer than one-second (Renshaw, 1945).

Vicory claims the WEFT system is superior to the Renshaw method based on a study that compared whether instructors referred only to the "total form" of an aircraft or "analysis of features" of an aircraft. One group used the Renshaw method and was not allowed to discuss features of the aircraft, while the other group did not use a tachistoscope and was able to discuss the features of the aircraft (Gibson, 1947; Vicory, 1968). This is an absurd argument to make against the Renshaw method. If an instructor is constrained from describing the features of an aircraft, it makes logical sense that this method would not enhance a cadet's ability to properly recognize an aircraft. There is no logical reason to use the WEFT system to focus on an "analysis of features" and the Renshaw system to focus on the "total form" in isolation. Indeed, years after both systems were created, the British created the Sargent System (1956) and the George Washington Human Resources Research Office created the HumRRO system; both of which used parts of the WEFT and Renshaw methods in conjunction with each other (Vicory, 1968).

d. A Tool in Collaboration

The Tachistoscope is not a tool required to be used in isolation, it should be part of a larger program or curriculum. The Renshaw method has been shown to be effective in recognizing different aircraft (Gibson, 1947; Renshaw, 1945). However, it has also been shown that providing a description of the aircraft and doing an "analysis of features" is also helpful. It is conjectured that the WEFT system would aid in distinguishing between very similar aircraft (e.g., F-15C vs. F-15E). While some aircraft look similar, their roles and capabilities could be very different. This topic will require evaluation in future studies and is outside the scope of this study.

C. TRAINING INTERVALS AND SPACING

1. Massed Practice versus Spaced Practice

The design of this study will utilize spaced practice vice massed practice. Over the course of two-weeks, participants will participate in about 180 minutes of training. This method was chosen because spaced practice has been shown to produce better effects than massed practice. A meta-analysis of 63 studies showed individuals in spaced training performed better than those who performed massed practice by almost 0.5 standard deviations (Donovan & Radosevich, 1999).

The nature of the task did play a role in the effects of massed versus spaced practice. Spaced training was particularly effective for tasks that that were lower in complexity and mental requirements (Donovan & Radosevich, 1999). Interestingly, when length of intertrial times were compared, stronger effects were shown for shorter intertrial times. This lends to the idea that learners do require rest, but not too much rest between tasks (Donovan & Radosevich, 1999). When working with cognitively demanding memory tasks, it has been shown that frequent training with short intervals is beneficial (Kennedy et al., 2015). Based on these findings, this study will entail extended training for two-weeks with a total of eight training sessions for 15–30 minutes, vice a single session used to achieve mastery. A longer period of study would be preferred. However, due to logistical considerations, the length of this study will be pared down in attempt to lower participant attrition.

2. Intervals between Images

The blank interstimulus interval (ISI) has been shown to be a component in an individual's ability to remember images. Memory performance increased from 20% with no ISI to 84% correct identification with an ISI of 1.5 seconds (Potter & Fox, 2009). Therefore, the tachistoscope should have an ISI of 1.5 seconds.

D. SUMMARY

Over the past 100 years, the tachistoscope has been used to address many problems. The tachistoscope combined with the RTS have shown success in multiple studies. However, RTS pairs well with other methods, such as the WEFT system. Recognition-Primed Decision Making is a proven model that works well in many different situations. Tachistoscopes support implicit learning and are an effective tool in the development of Recognition-Primed decision making. It is part of Gary Klein's Naturalistic Decision-Making model that all people use on a regular basis.

THIS PAGE INTENTIONALLY LEFT BLANK

III. RETENTION SOFTWARE PROGRAM

A. OVERVIEW

The Retention Software Program, known as Retention, was designed by Major Brian T. Pugh, and built by the Naval Postgraduate School's Futures Tech Team in Unity. The program was built to be simple and intuitive for users. All settings were preset by the research team. Users would login and select the current test or training period to begin their session. Two separate versions were built, one for a PC and one for an Oculus VR headset. A third version was being created for an HTC headset, but development stopped because it was no longer needed. The requirements document provided to the NPS Futures Tech Team is included in Appendix A.

B. DESIGN

To test the hypotheses in this thesis a system was needed to allow researchers to build fully customizable training and testing sessions. The researchers needed to be able to change the objects, images assigned to objects, how many times the images were presented, and how long those images are presented. For this research, a user interface was not created. All the required adjustments were done through the manipulation of json files.

1. Overview

The user will log into the system using a provided username and password. If this is the first time they logged in, they will go through a short introduction. Otherwise, the user will skip the introduction and choose the available training or testing session. The overall process for an entire session is shown in Figure 9.

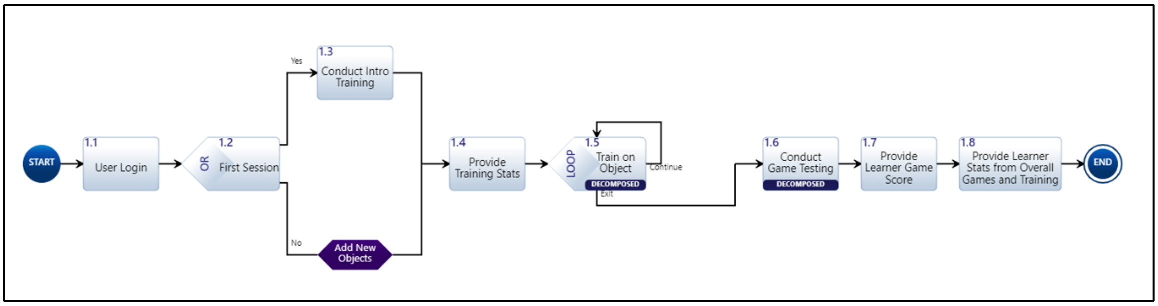


Figure 9. Tachistoscope Operation Action Diagram

When a user opens the program, they will come to the log in screen. The first series of screens in the program are seen in Figure 10.

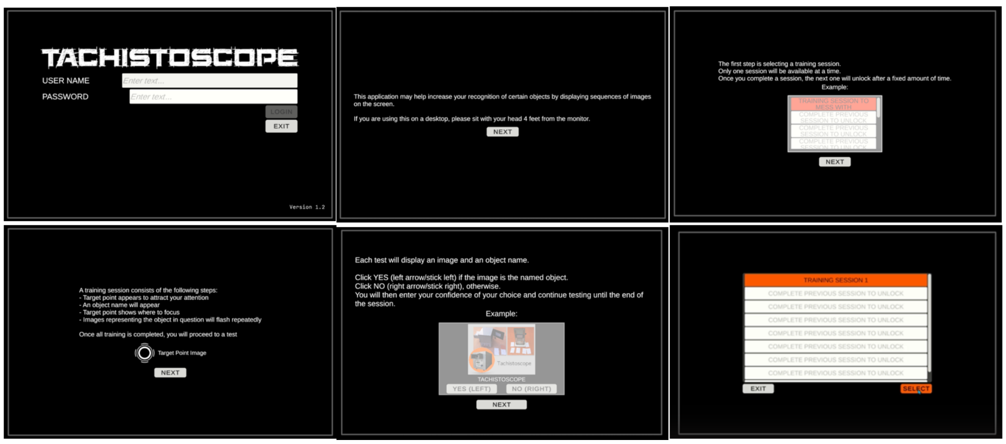


Figure 10. Sample of Login Screen and Training Session Selection Screen

2. Training Sessions

The action diagram in Figure 11 is an expanded view of Block 1.5: Train on Object in Figure 9. The action diagram in Figure 11 will be iterated for each object in the training session. The researcher or trainer can adjust the name of the object, which images are presented, how many times the images will be presented, and how long the images will be presented.

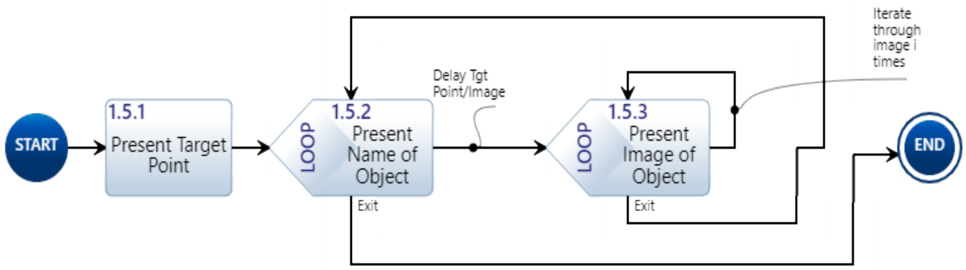


Figure 11. Training Session Action Diagram

Figure 12 provides an example training session for two objects. The aircraft presented in the training session were ordered randomly within their category of aircraft (e.g., fighter, bomber, commercial airliner, etc.).

Figure 12. Example of Training Session with Two Objects
Adapted from *J-20 - Fighter Aircraft in Flight* (2011); Turner (2022).

3. Immediate Testing

Immediately after a training session is complete, the user will receive testing on the objects observed during that particular training session. The process for the immediate testing portion of the program is shown in Figure 13. The user will receive brief instructions on how to take the test and then the program will iterate through each object as seen in block 1.6.4.

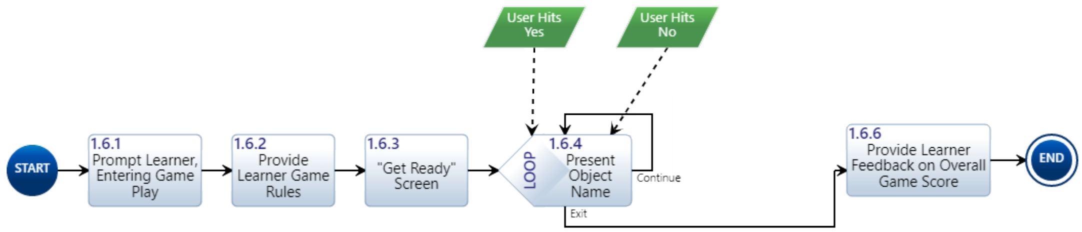


Figure 13. Testing Action Diagram

Figure 14 provides an example of a test question. Response Time is measured from the moment the question is presented until the user chooses their answer. The next screen asks the user to provide their level of confidence on that question. The confidence level is a sliding scale from 0 to 100.

Figure 14. Example of Test Question. Adapted from *Antonov 225 - Russian Transport Aircraft* (2007).

After completing the immediate testing, the user will receive overall feedback on the number correct answers. Once the user concludes their immediate testing the csv file is written and sent to the Amazon Web Service (AWS) cloud.

C. SCRIPTS

1. Scripts Required to Run Retention Software Program System

The scripts listed below fed into the system to direct the training sessions, the order of the training sessions, which images to draw from for which objects, timing, and user information.

- Application Configuration.json. Trainer can create a time gap between training sessions and the presentation speed of images.
- Images.json. A list of dictionaries that pairs the name of the image to an asset path for the image.
- Timing Configuration.json. Trainer can create different timing configurations for training.
- Training Sessions.json. A list of the training sessions. The order determines the order users will proceed through training.
- User Accounts.json. A list of dictionaries for each user that includes username, password, and their timing configuration.
- The individual Training and Testing Session json files. A list of dictionaries that determines the order in which objects and their images are presented. It also includes a list that determines test objects, their image, and if the object name matches the image.

2. Scripts Required to Build the Scripts in Paragraph III.C.1

The scripts below were used to write the scripts needed to use the Retention Software Program System.

- build_training_session.py. This script builds all the training sessions. It will read image names from the folder with the images and then put the images in categories. Trainers can create a session name, determine the number of aircraft per category that will be used in that session, the number of images per object, and the presentation timing. The selection of aircraft and images is random. The order in which they are presented is random.

- create_images.py. This script reads the directory will all of the images and creates the Images.json file based upon the file names.
- convert_username_excel_to_json.py. This script allowed the researchers to take an excel spreadsheet with usernames, passwords, and timing and convert the data to a json file, User Accounts.json.

IV. METHOD

A. OVERVIEW

The study consisted of a controlled experiment with a between group pre-post design. There were two experimental groups and one control group (Experimental groups: Tachistoscope on VR Headset and Tachistoscope on Laptop groups). All participants, regardless of group, completed the same computerized pretest to capture each participant's current recognition and knowledge of the aircrafts being tested. The same test format, with different images, was used for the post-test.

The experiment was conducted over two and a half weeks in which participants were trained to recognize (40) different aircraft over eight training sessions conducted Monday to Friday each week. The pre-test included all (40) aircraft and was conducted throughout the week prior to the training sessions. The first training session had (10) aircraft. Each subsequent training session added five additional aircraft. The seventh and eighth training sessions both included all (40) aircraft. Following each training session, immediate testing that included only the aircraft trained on that day was conducted. The post-test included all (40) aircraft and was scheduled to be conducted the Monday following the eighth training session. Due to scheduling issues, two of the participants (110 and 225) had to take the posttest the Friday immediately following the last training session.

B. PARTICIPANTS

1. Selection

The Naval Postgraduate Institutional Review Board approved the participant solicitation and all research methods used in this study (approval number NPS.2022.0017-IR-EP7-A). Participants were treated in accordance with Department of Navy standards. Participants were all informed of their rights and signed a consent form prior to any surveys or experimentation. Due to the flashing nature of a tachistoscope participants were asked if they had visually induced epilepsy, none answered yes.

Participants were randomly assigned to one of the three groups. The study started with 24 participants; one participant had to drop from the study due to personal issues unrelated to the study. Initially each group had eight participants. Due to scheduling issues, one of the control group participants had to be moved to the Tachistoscope on VR Headset group.

C. TEST AND TRAINING TASKS

The data set to train and test the participants included 40 aircraft separated into five categories, as seen in Table 1. A mix of U.S., European, Chinese, and Russian aircraft were chosen. Ten different images were selected for each aircraft. The images included different angles and perspectives and aircraft with different paint schemes to the greatest extent possible. Scripts were used to write the json files needed for the program to run the training sessions. More detail on the different scripts required can be found in Chapter III Retention Software Program.

Table 1. Aircraft Listed by Category

Fighter	**Business Jet**	**Commercial Airliner**	**Bomber**	**Cargo Military**
F-15C	G600	787	B-1	C-2
F-35	B200	A380	B-2	C-5
T-38	CJ1	CRJ1000	B-52	C-17
Su-57	DF900	737	F-117	C-130
J-10	BAE-146	DC-10	H-6K	An-140
J-20	C172	E190	Tu-95	Il-76
Mirage-2000	HondaJet	747	Tu-22M-3	An-225
GR-4-Tornado	SR-22	B1900	Tu-160	An-26

1. Pretest

The pretest included 40 questions, one for each aircraft. Each question was randomly assigned to display an image that either matched or did not match the aircraft title below the image. The title of the object matched the image in about half of the questions. If the question was assigned to have the image match the text, the image was chosen from one of that aircraft's ten images. If the question was randomly selected to have the image not match the text, the image was randomly selected from a different aircraft in the same category. Figure 15 is an example of a question in which the image does not match the aircraft title. For this question, the participant should select the "No" button Figure 15. The participant's response time was measured from the moment the image is presented to the participant, until the participant clicked the button.

Figure 15. Example Test Question. Adapted from Axel (n.d.).

2. Training Tasks

There were eight training sessions, further described in the Procedures section. Each session followed the same pattern. During the training sessions the name of the aircraft was presented for one second after which the aircraft image was presented. This loop occurred for each image. Each aircraft had six images randomly selected from its set of ten, that were presented in succession. Once all images were presented, a screen appeared with a box the participant clicked if they were ready to proceed to the next aircraft. An example of a training image, F-15C, is shown in Figure 16. The left box would be presented first and then the image on the right. All participants in all groups viewed the same sequence of aircraft and images. The only differences between the groups were the time the image was presented and the frequency in which the image was presented. Images in the Tachistoscope Groups were presented for 250 ms five times with an ISI of 500 ms. Images in the Control Group were presented for five seconds one time.

Figure 16. Training Image Example F-15C. Adapted from Turner (2022).

At the end of each training session, all participants took part in immediate testing. The immediate test was the same for all participants and only included the aircraft that were presented on in that training session. The test was generated in the same manner as the pretest and posttest but was taken on the participants' respective training device (VR headset or Laptop).

3. Posttest

The posttest was in the same format as the pretest.

D. MATERIALS

1. Equipment

a. Two Oculus Quest 2

- Processor: Qualcomm® Snapdragon XR2 Platform
- Resolution: 1832 x 1920 at 72 hertz per eye
- Tracking: 6 degrees of freedom
- Memory: 6 GB

b. Four Laptop Computers - MSI GE66 Raider

- Processor: 11^{th} Gen Intel(R) i9-11980HK @2.60GHz
- Resolution: 3840 x 2160 at 60 hertz

- 15.6” Screen
- Installed RAM: 32 GB
- Windows 10 64-bit Operating System
- NVIDIA GeForce RTX 3080 Laptop GPU

E. VARIABLES

1. Independent Variables

Three independent variables were collected: presentation time, frequency of image presentation, and type of training device.

a. Presentation Method: Time and Frequency

The number of times each image was presented and the frequency at which each image was presented varied between the two Tachistoscope Groups (VR Headset and Laptop) and the Control Group. In the Tachistoscope Groups, images were presented five times at a frequency of 250 milliseconds. Between those presentations there was a 500 millisecond ISI in which the screen was completely black. The Control Group viewed each image one time for five seconds.

b. Type of Training Device

There were two types of training devices used for this experiment. The Control Group and the Tachistoscope on the Laptop Group both used the same laptops. The Tachistoscope on the VR Headset Group used an Oculus Quest II virtual reality headset.

2. Dependent Variable

Two dependent variables were examined, difference in pretest to posttest accuracy and difference in pretest to posttest mean response time.

a. Number of Correct Identifications

The Tachistoscope program was able to capture the answers for all the participants' answers throughout the study. Each session saved the answers in a separate csv file. For each participant, the difference from pretest to posttest was measured by ($Posttest_{total}$ - $Pretest_{total}$).

b. Reaction Time to Answer Question

As with the participants' answers, the Tachistoscope program was able to capture the reaction time for each question. Once the image was presented, a timer captured how long it took for the participant to answer "Yes" or "No." The participants were not told their reaction time was being measured. For each participant, the difference in reaction time was measured by ($Posttest_{RT}$ – $Pretest_{RT}$).

3. Output Files

At the conclusion of the training session or testing period the Retention Software Program generated a csv file that include the participants name, training session, and the time they logged into the system. In addition, the Tachistoscope system captured the aircraft name (Test Started), the answer the user submitted (Test Submitted Answer), if the answer was correct (Test Graded), the time it took the participant to answer once the image appeared (Time on Test (in seconds), and the participant's confidence level for that particular question (User Confidence (0 – 100).

Table 2. Example Participant Data File

0.0029993	start-time	12:02:00
0.0039998	User Logged In	User100
0.0137284	Tutorial Started	
4.4993981	Tutorial Step Completed	1
19.2976106	Tutorial Step Completed	2
36.2139314	Tutorial Step Completed	3
49.7146304	Tutorial Step Completed	4
51.364688	Tutorial Step Completed	5
51.3666962	Tutorial Completed	
53.7639362	Training Session Started	Pre-Test
53.7639362	Training Started	
53.8142618	Training Completed	
58.5988282	Test Started	BAE-146
70.8985605	Test Submitted Answer	No
70.8985605	Test Graded	Correct
70.8985605	Time On Test (in seconds)	12.29973
74.5647463	User Confidence Level (0-100)	46.67
74.5647463	Test Started	HondaJet
78.5977771	Test Submitted Answer	No
78.5977771	Test Graded	Correct
78.5977771	Time On Test (in seconds)	4.033031
80.097596	User Confidence Level (0-100)	50

F. PROCEDURE

Prior to the conduct of the experiment, all participants completed a survey to track demographics and experience level. The week prior to training, participants came into the lab and the researcher explained the procedure of the study to the participants and the participants completed a pretest.

Table 3. Week-By-Week Activities

Day of the week	**Week 1 Preparatory activities**	**Week 2 Training Sessions 1–4**	**Week 3 Training Sessions 5- 8**	**Week 4 Post training activities**
Monday Tuesday Wed Thurs	Consent Demographic survey Pretest	Sess 1: 10 aircraft Sess 2: 15 aircraft Sess 3: 20 aircraft Sess 4: 25 aircraft	Sess 5: 30 aircraft Sess 6: 35 aircraft Sess 7: 40 aircraft Sess 8: 40 aircraft	Posttest Posttest Survey

Following the pretest, each participant executed eight training sessions, one session per day Monday to Thursday for two consecutive weeks. All subjects in all groups were required to participate in 15 – 30 minutes of training immediately followed by about five minutes test. In the first training session, participants were trained on 10 aircraft. For each successive training day, the number of aircraft were increased by five, such that by training session 7, participants were trained on all 40 aircraft. (see Table 3). Both the control group and the experimental groups saw the same aircraft and images during their respective training sessions and completed the same immediate test after each training session on their respective training devices.

For the control group, the name of the aircraft was presented for one second and then the image was shown once for five seconds.

For the Tachistoscope groups, the images of the aircraft were presented to the participant five times at a rate of 250 milliseconds with a blank interstimulus interval of 500 milliseconds (Potter & Fox, 2009). Once the set of six images was complete, the next aircraft was presented in the same manner. The only difference between the Tscope VR Headset and Tscope Laptop groups were the platform device on which the training occurred.

At the end of each training day the research team pulled the scores for each of the participants. The participants were then ranked from 1 to 23 based on their score. A leaderboard of the top 10 participants, identified solely by participant id, was sent out to participants to maintain motivation and interest in the study.

Once the training session 8 was completed, each participant took a posttest on the laptop. Participants were scheduled to take the posttest the Monday following the last training session on Thursday. The delay was intentional to see allow time between training and testing. Unfortunately, two participants had to take the posttest the day after the last training day due to scheduling conflicts, participants 110 and 225. Both were within one standard deviation (2.29) from the mean (36.09) and showed no significant advantage.

THIS PAGE INTENTIONALLY LEFT BLANK

V. RESULTS

This chapter will explain the results of the study. First, an analysis of the Participants will be discussed, to include the demographics for each of the groups. The next section will be an overview of the statistical methods used to analyze the hypotheses. Then the descriptive statistics for the Scores, Reaction Times, and Confidence Levels will be presented. This section will lead into the analyses of Hypothesis 1 and 2, which will be followed by a series of exploratory topics to provide more depth to the study. Finally, the Chapter will be concluded with General Observations on the study.

A. PARTICIPANT ANALYSIS

The study began with 24 participants and finished with 23. One participant could not complete due to personal reasons.

As shown in Table 4, the three groups did not vary in demographic variables, including the variables, "Previous Work With Aircraft" ($p > .75$) and "Received Aircraft Identification Training," ($p > 0.77$).

Table 4. Distribution of Participants

	Tscope VR Headset (n = 9)	**Tscope Laptop (n = 8)**	**Control (n = 6)**
Mean Age (sd)	34.1 (6.2)	31.8 (4.7)	33.7 (3.8)
Service Branch (frequency)	(3) Army, (1) Navy, (5) USMC	(2) Air Force, (1) Coast Guard, (1) Navy, (4) USMC	(1) Air Force, (2) Army, (3) USMC
Mean Years of Service (sd)	12.4 (5.6)	9.6 (4.9)	10.3 (4.6)
Functional Area (frequency)	(3) Aviation, (1) Fires, (1) Logistics, (1) Maneuver, (2) Other	(1) Aviation, (2) Logistics, (2) Maneuver, (3) Other	(1) Fires, (1) Logistics, (3) Other
Gender (Male/ Female)	9 Male / 0 Female	7 Male / 1 Female	6 Male/ 1 Female

	Tscope VR Headset (n = 9)	**Tscope Laptop (n = 8)**	**Control (n = 6)**
Highest Rank	(1) O-1, (3) O-3, (3) O-4, (2) O-5	(1) O-1, (3) O-3, (4) O-4	(1) O-1, (3) O-3, (2) O-4
Previous Work with Aircraft (% Yes)	44%	63%	50%
Received Aircraft Identification Training (% Yes)	22%	38%	33%
Used VR device before (% Yes)	67%	88%	50%

B. STATISTICAL ANALYSIS METHODS

Descriptive statistics were run on the Pretest and the Posttest Scores and Reaction Times to get an overall sense of the data. Scores and Reactions Time were further broken down into Overall, the Control Group, and the two Experimental Groups.

To test Hypothesis 1, the researchers took the difference in Score within each participant from Pretest to Posttest and then compared the mean difference scores between groups using an ANOVA test. To test Hypothesis 2, the researchers took the difference in Mean Reaction Time within each participant from Pretest to Posttest and then compared the Mean Reaction Time between groups using an ANOVA test. Regarding Pretest Score there was one outlier who was in the Control Group, Participant 325. Their Pretest Score was 35 out of 40 due to prior extensive knowledge of aircraft. Hypothesis 1 will be analyzed with and without Participant 325 included. Both hypotheses will use an alpha level of 0.05.

C. DESCRIPTIVE STATISTICS

1. Pretest

Overall, participants took about 5–10 minutes to complete the test. The mean score for all participants was 25.3 out of 40 correct answers with a standard deviation of 5.0

correct answers, as seen in Figure 17, and an average participant confidence level of 42 out of 100.

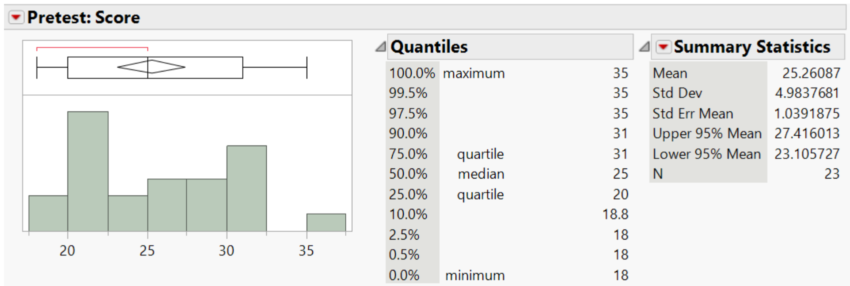


Figure 17. Overall Pretest Score Descriptive Statistics

The mean Reaction Time for all participants was 5.86 seconds with a standard deviation of 1.79 seconds, as seen in Figure 18. Reaction Time data was normal except for one outlier.

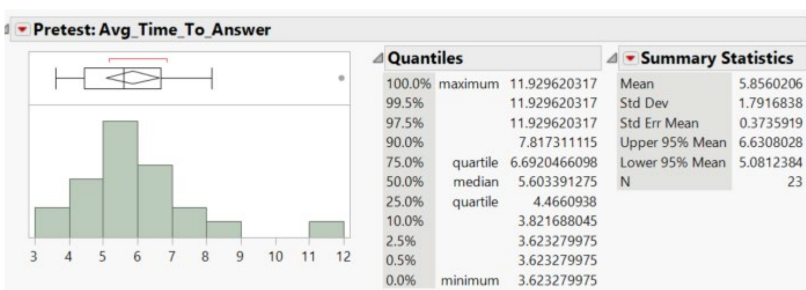


Figure 18. Overall Mean Reaction Time Descriptive Statistics

a. Control Group Pretest

Control Group scores were not normally distributed. Five out of six scores were either 20 or 25. There was one outlier (participant 325) who had 35 correct answers. This left little room for 325 to improve over the course of the training. The Reaction Times were evenly distributed from 3.6 to 7.0 seconds, with a mean of 5.6 (sd = 1.4).

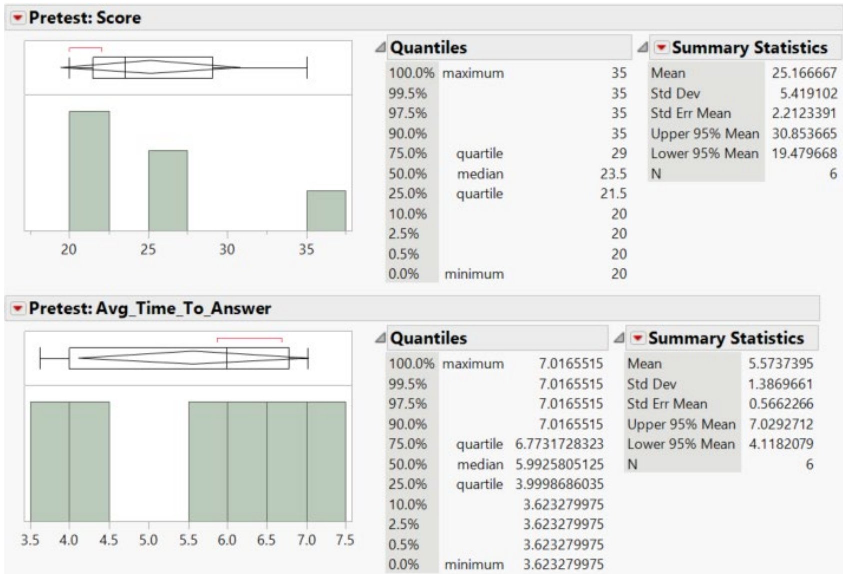


Figure 19. Control Group Pretest Scores and Mean Reaction Time

b. Tscope on Laptop Group Pretest

The Tscope on Laptop Group scores had a wide distributed range, with scores ranging from 18–31, with a mean of 24.38 (sd = 5.5). The Reaction Time was mostly normal with a range of 3.7 to 6.4 seconds, with a mean of 5.0 seconds (sd = 0.9 seconds).

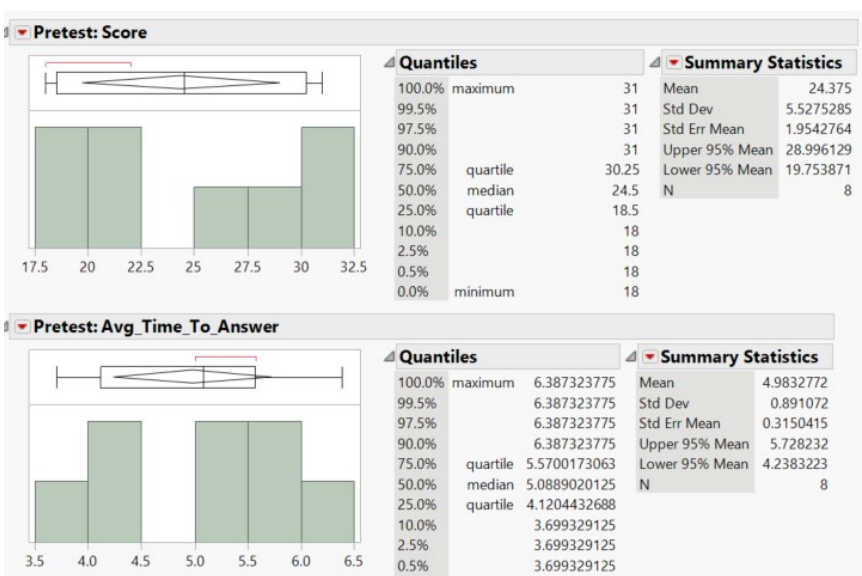


Figure 20. Tscope on Laptop Pretest Scores and Mean Reaction Time

c. *Tscope on VR Headset Pretest*

The Tscope on VR Headset group had a mean test score slightly higher than the other two groups, but it was well within one standard deviation of both other groups. The scores were not normally distributed and there were two groupings of scores, one between 20–24 and another between 28–32. Reaction Time is somewhat normal and skewed right, with a mean of 6.8 seconds (sd = 2.25), as seen in Figure 21. There was a trend for Mean Reaction Time to be slower than both other groups, F (2, 20) = 2.68, , p = 0.093.

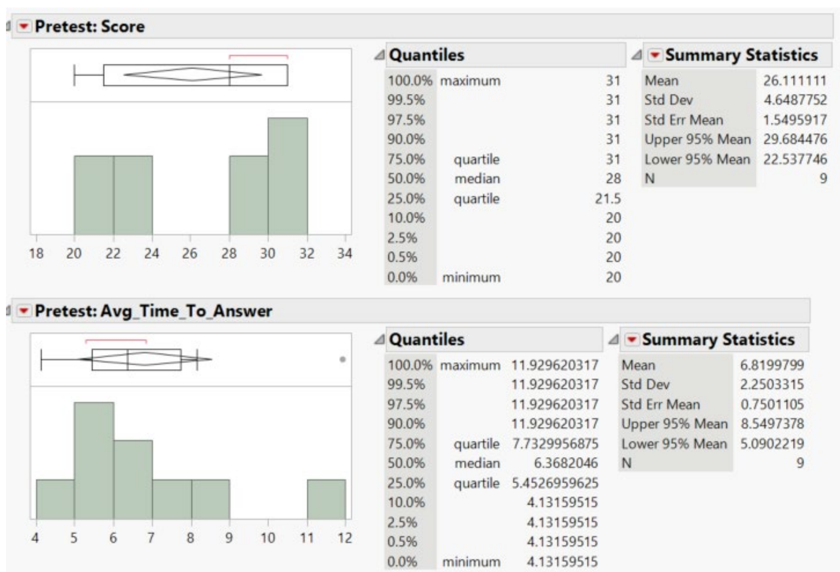


Figure 21. Tscope on VR Headset Pretest Scores and Mean Reaction Time

2. Posttest

The mean Score for the Posttest for all participants was 36.0 with a standard deviation of 2.3. The scores were moderately normal. The Reaction Time was somewhat normal, with a mean of 5.5 seconds (sd = 1.7 seconds)

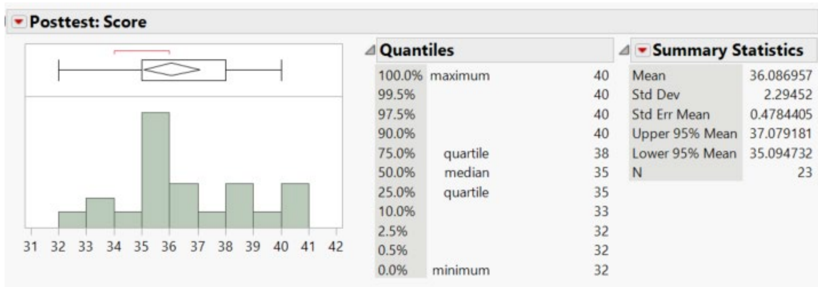


Figure 22. Overall Posttest Descriptive Statistics for Scores

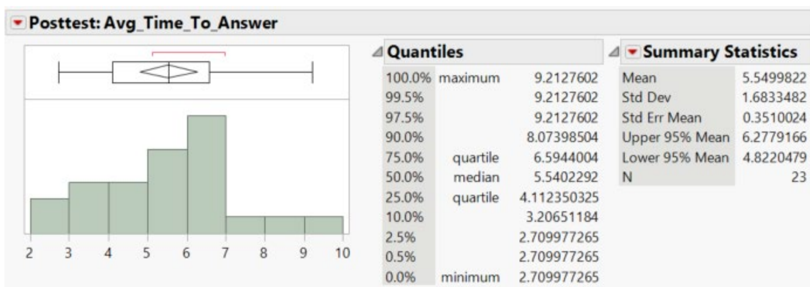


Figure 23. Overall Posttest Descriptive Statistics for Reation Time

a. Control Group Posttest

The mean score for the Control Group was 36.5, (sd = 1.9). The Scores for the Control Group are mostly normal. The Mean Reaction Time was spread across a wide range, 5.86 seconds; the majority of participants averaged 5–6 seconds per question. The mean Reaction Time was 5.18 seconds, (sd = 2.0 seconds).

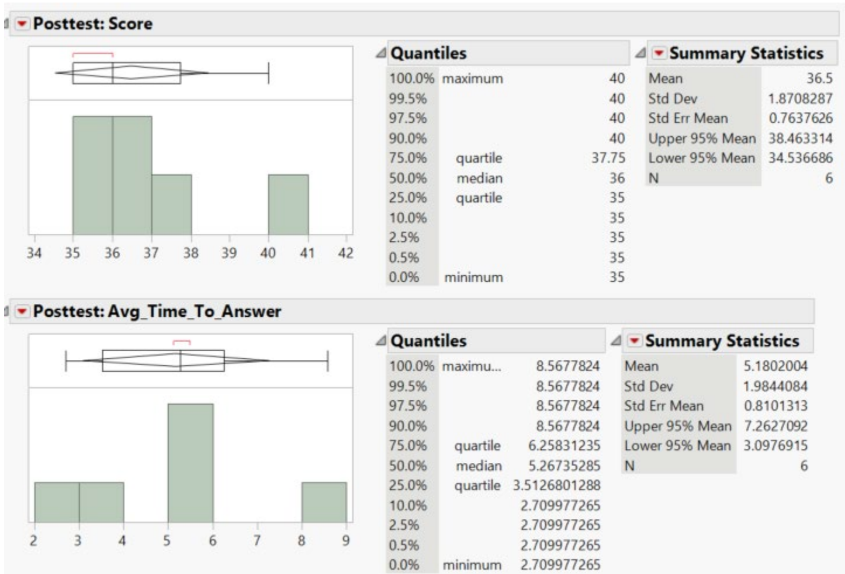


Figure 24. Control Group Posttest Score and Mean Reaction Time

b. Tscope on Laptop Group Posttest

The Scores for the Tscope on Laptop Group were slightly normal and had the largest range of the three groups. The mean Score was 35.9, (sd = 2.7). The Reaction Times had a wide range, 4.4 seconds, and were spread evenly across that range. The mean Reaction Time was 5.1 seconds, (sd = 1.4 seconds).

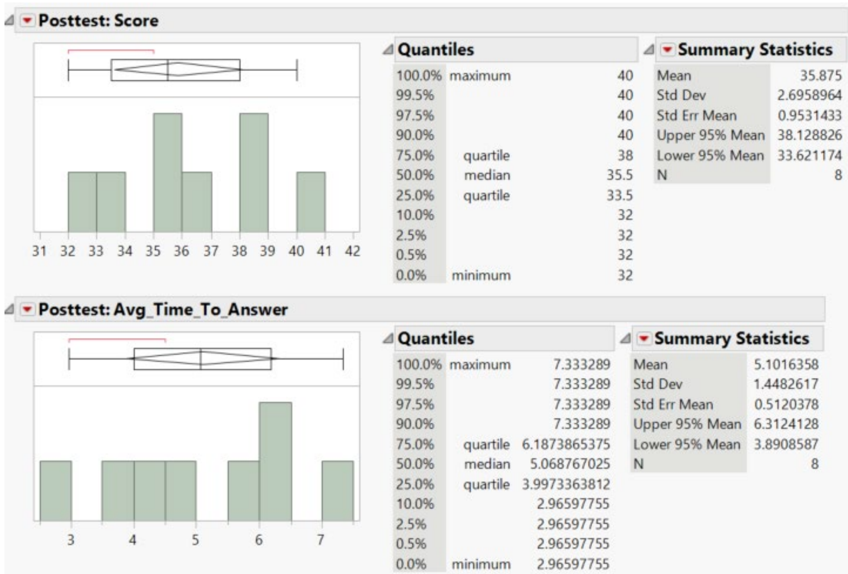


Figure 25. Tscope on Laptop Group Posttest Score and Mean Reaction Time

c. Tscope on VR Headset Group Posttest

The Scores for the Tscope on VR Headset Group were separated into two groups, (33-36) and (38-41) with a mode Score of 35. The mean Score was 36 with a standard deviation of 2.4.

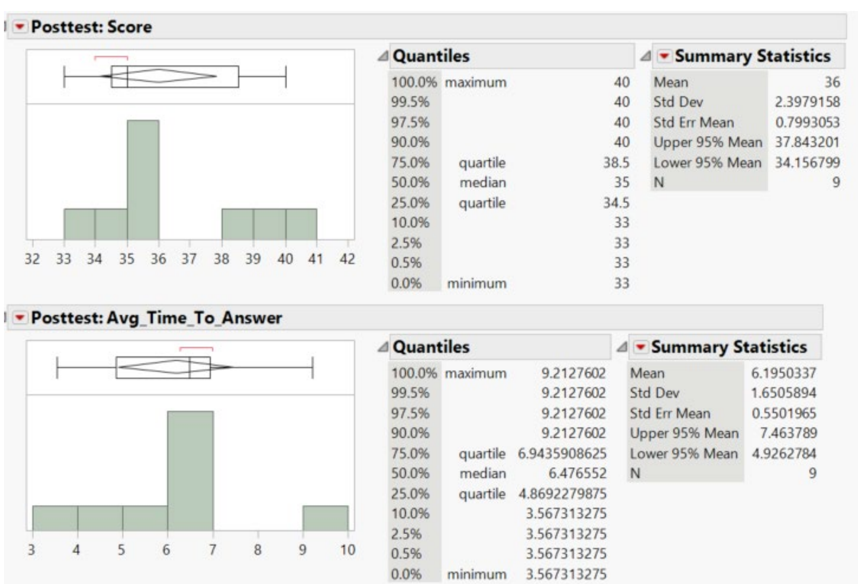


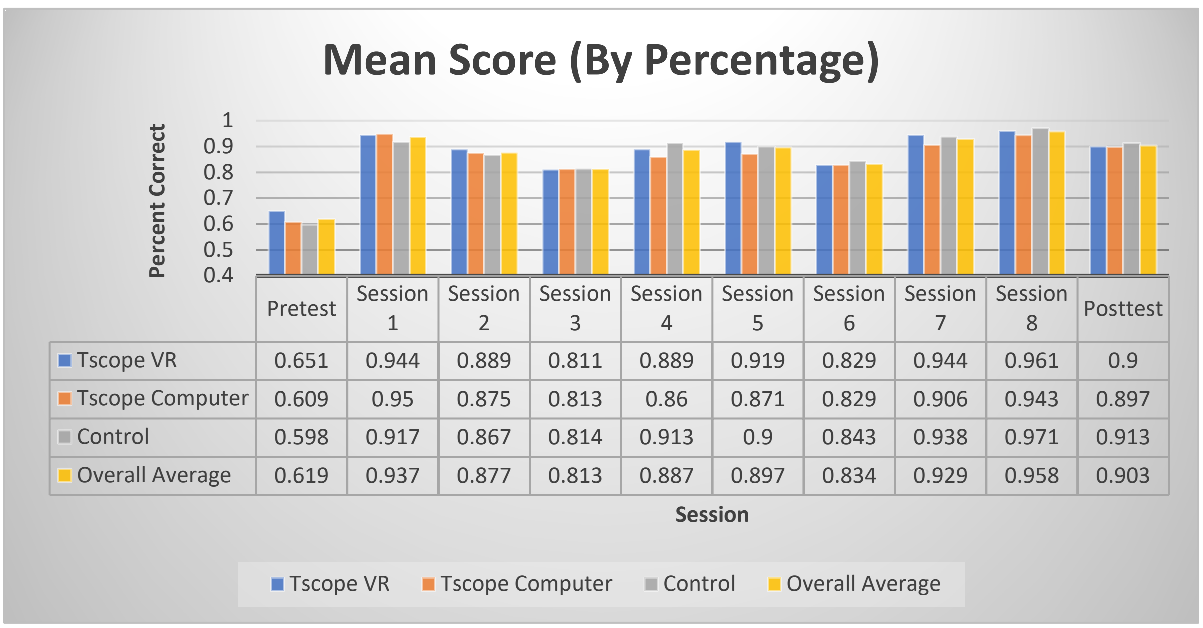
Figure 26. Tscope on VR Headset Group Posttest Score and Average Mean Reaction Time

3. Mean Difference in Score and Average Reaction Time by Session

The results of the Pretest, Posttest, and the quiz that followed each training session are captured in Figure 27 and Figure 28. It is important to note, not all training sessions had the same number of aircraft. Training Session 1 had 10; each subsequent training session added five until all 40 aircraft were included.

a. Improvement in Test Scores over Training Sessions

In general, the Scores improved as training progressed from Pretest to Training Session 8. From Training Session 8 to Posttest there was a slight decline, as seen in Figure 27.

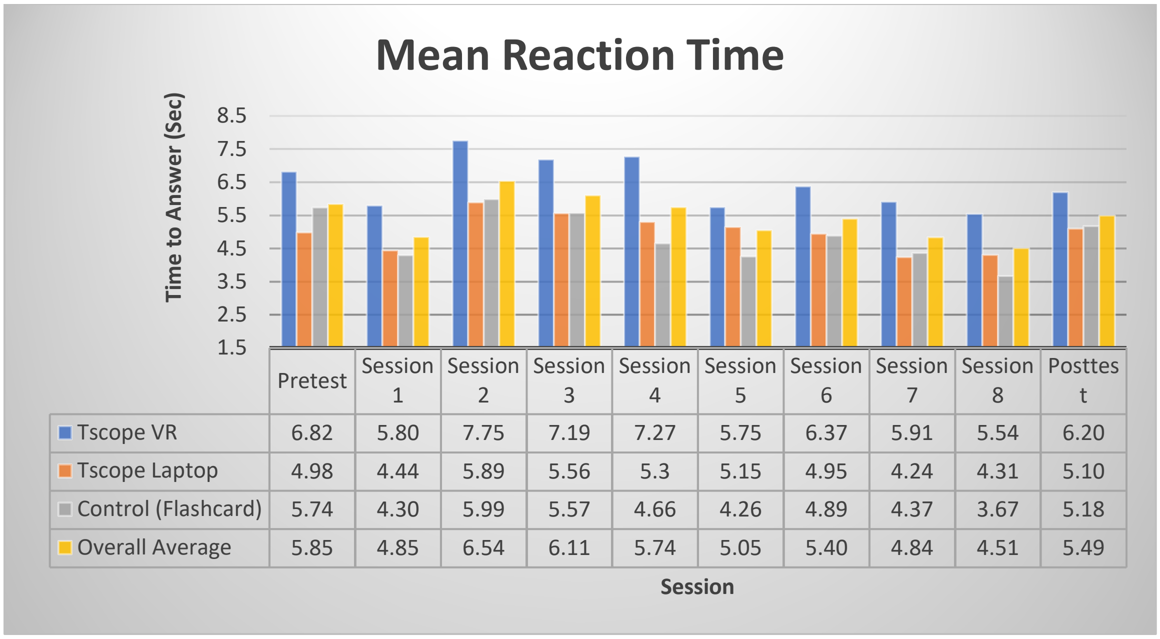


	Pretest	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8	Posttest
Tscope VR	0.651	0.944	0.889	0.811	0.889	0.919	0.829	0.944	0.961	0.9
Tscope Computer	0.609	0.95	0.875	0.813	0.86	0.871	0.829	0.906	0.943	0.897
Control	0.598	0.917	0.867	0.814	0.913	0.9	0.843	0.938	0.971	0.913
Overall Average	0.619	0.937	0.877	0.813	0.887	0.897	0.834	0.929	0.958	0.903

Figure 27. Mean Score (by percentage) for Each Session Graph and Table

b. Reaction Time Varied across Training Sessions

In general, the Reaction Times became quicker from Pretest to Training Session 8. From Training Session 8 to Posttest Reaction Time became slightly slower. There was a slight increase in Reaction Time from Session 8 to the Posttest, further analysis is in paragraph F.4.



	Pretest	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8	Posttest
Tscope VR	6.82	5.80	7.75	7.19	7.27	5.75	6.37	5.91	5.54	6.20
Tscope Laptop	4.98	4.44	5.89	5.56	5.3	5.15	4.95	4.24	4.31	5.10
Control (Flashcard)	5.74	4.30	5.99	5.57	4.66	4.26	4.89	4.37	3.67	5.18
Overall Average	5.85	4.85	6.54	6.11	5.74	5.05	5.40	4.84	4.51	5.49

Figure 28. Mean Reaction Time for Each Session Graph and Table

D. HYPOTHESIS 1: TESTING GROUP DIFFERENCES IN PRE – POST TEST SCORES

H_{O1}: There is no difference between the standard method group (control), computer Tscope group, and the Tscope on the VR platform group in the average increase in the number of objects identified before and after training. $\mu_{control\text{-}d} = \mu_{VRTscope\text{-}d} = \mu_{CompTscope\text{-}d}$.

H_{A1}: There is a difference between the standard method group (control), computer Tscope group, and the Tscope on the VR platform group in the average increase in the number of objects identified before and after training. $\mu_{control\text{-}d} \neq \mu_{VRTscope\text{-}d} \neq \mu_{CompTscope\text{-}d}$.

While the within participant Scores increased in each group, there was only a slight variance in mean Scores between groups. The p-value for the overall model was greater than 0.74; the model is not statistically significant. Thus, the null hypothesis cannot be rejected; there is no significant difference in mean Scores between groups.

The Pretest had the lowest Scores. Across the eight training sessions, test scores ebbed and flowed as the participants balanced their knowledge of the aircraft and the number of aircraft included in each immediate test. Overall, the average Scores trended in a positive direction. The highest average Score was in Session 8, which had all 40 aircraft in the immediate test. Session 8 also had eight participants with perfect Scores. Since the Posttest was conducted 4-days after the last training session, it stands to reason there would be memory decay; from Session 8 to the Posttest, the average overall Score decreased from 96% correct to 90% correct.

Table 5. Mean Difference in Scores (Post – Pre)

Control Group	Tscope on Laptop	Tscope on VR Headset
11.3	11.5	9.9

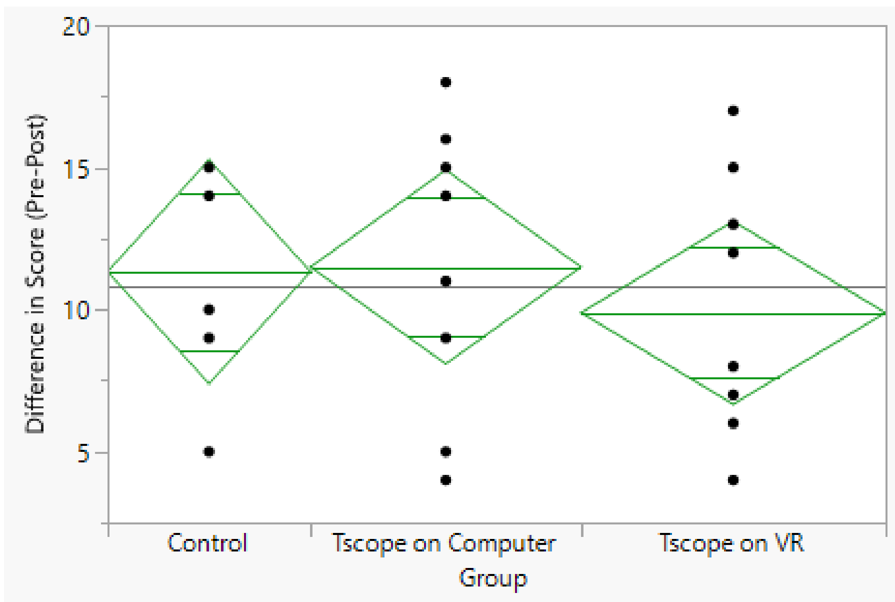


Figure 29. One-way Analysis of Difference in Score (Pre-Post) by Group

Table 6. Hypothesis 1: Difference in Scores ANOVA Results

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Group	2	13.08213	6.5411	0.3055	0.7401
Error	20	428.22222	21.4111		
C. Total	22	441.30435			

a. Data without Outlier Participant 325

The mean difference in Score for the control group increased by 1.3 when the outlier is removed. Even with this change, the ANOVA results were not significant, F (2,19) = 0.64, p = 0.54.

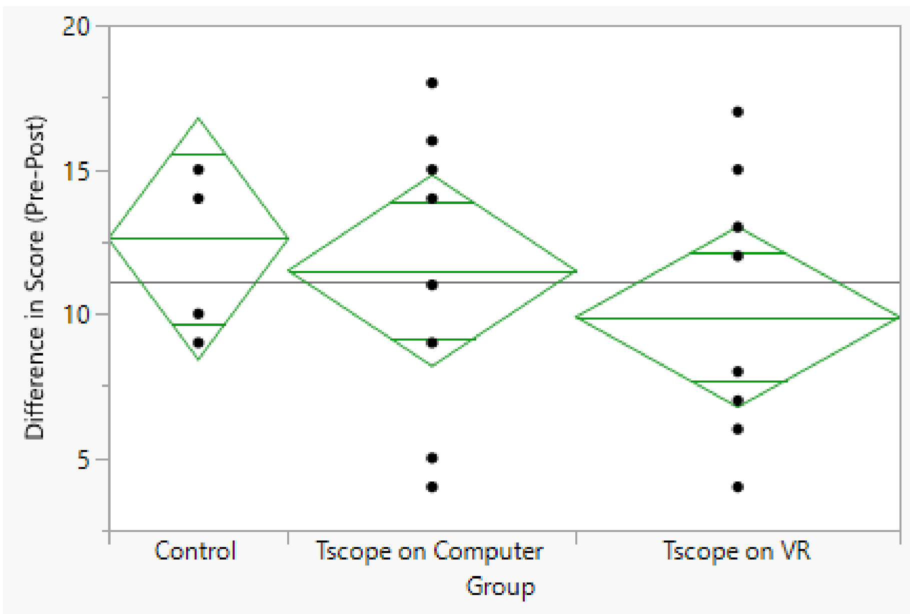


Figure 30. One-way Analysis of Difference in Score (Pre-Post) by Group without Outlier

b. Compare Results of Training Session 8 to Posttest, Pair T-Test

Paired t-tests were conducted to determine if there were overall differences in Scores and Reaction Times from Training Session 8 to the Posttest. An alpha level of 0.05 was used to conduct the paired t-tests

Across all participants, the mean difference in Score from Training Session 8 to Posttest is -2.3 with a standard error of 0.4. All but one Participant did better on Training Session 8 than the Posttest. The Participant who improved went from 39 correct to 40 correct. There is a significant difference between Training Session 8 and the Posttest, p-value < 0.0001. This would suggest there was some memory decay after the last day of training.

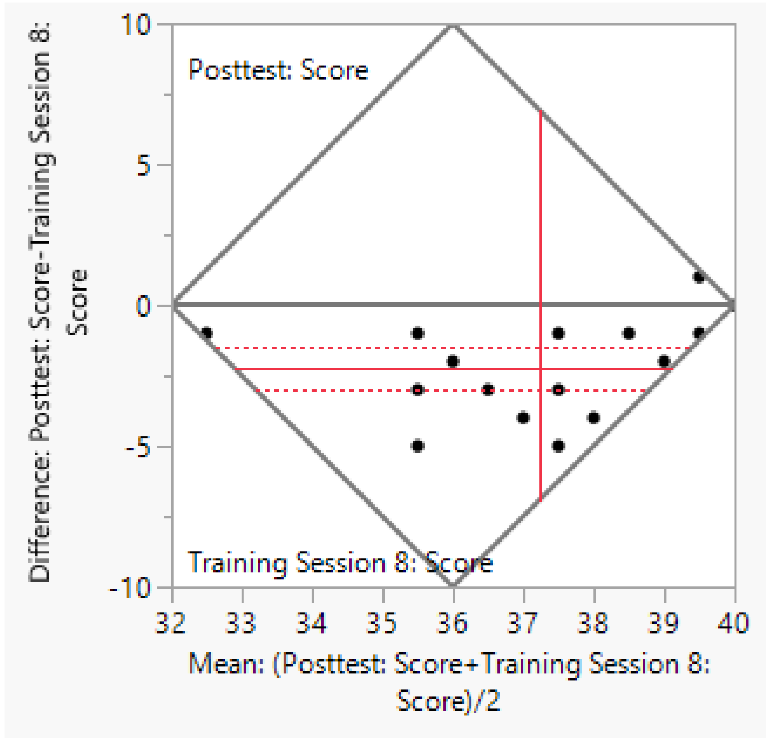


Figure 31. Matched Pairs: Difference in Score from Training Session 8 to Posttest Graph

E. HYPOTHESIS 2: TESTING GROUP DIFFERENCES IN PRE-POST CHANGES IN REACTION TIME

HO2: There is no difference between the Standard Method Group (Control), Tscope on Laptop Group, and the Tscope on the VR Headset Group in the average change in speed at which objects are correctly identified before and after training. $\mu_{dRT1} = \mu_{dRT2} = \mu_{dRT3}$

HA2: There is a difference between the Standard Method Group (Control), Tscope on Laptop Group, and the Tscope on the VR Headset Group in the average change in speed at which objects are correctly identified before and after training. $\mu_{dRT1} \neq \mu_{dRT2} \neq \mu_{dRT3}$

The mean Reaction Time difference from Pretest to Posttest is in Table 7; there is very little variance between groups. The p-value for the overall model was greater than 0.64; not statistically significant. Thus, the null hypothesis cannot be rejected; there is no difference in Reaction Time between groups. Results did not change when the Control Group outlier was removed.

Participants were not told Reaction Time was being measured until after the experiment. The overall average Reaction Time was the quickest in Session 8. Except for the Pretest and Session 1, overall average Reaction Time steadily became faster until Session 8.

Table 7. Mean Difference in Reaction Time (Pre – Post) (seconds)

Control Group	Tscope on Laptop	Tscope on VR Headset
0.62	-0.12	0.62

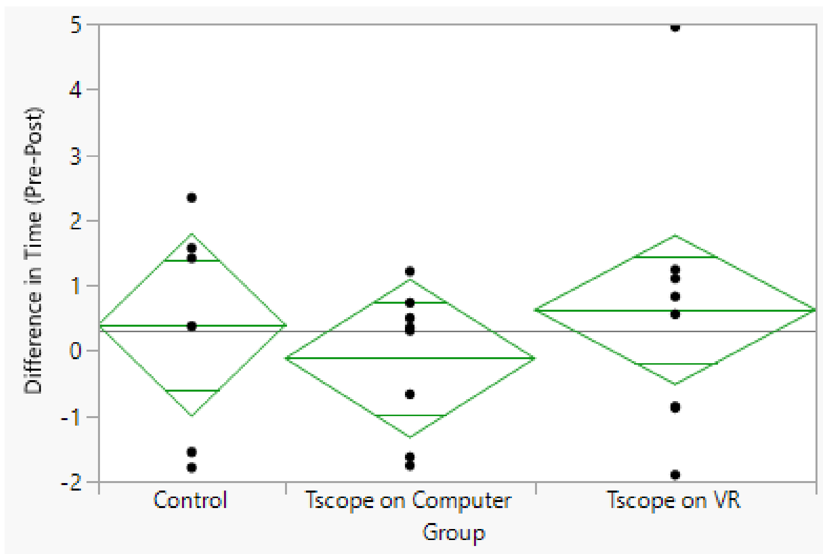


Figure 32. One-way Analysis of Difference in Time (Pre-Post) by Group

Table 8. ANOVA Difference in Time (Pre – Post) By Group

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Group	2	2.402161	1.20108	0.4485	0.6449
Error	20	53.560095	2.67800		
C. Total	22	55.962255			

a. Compare Results of Training Session 8 to Posttest, Pair T-Test

Paired t-tests were conducted to determine if there were overall differences in Scores and Reaction Times from Training Session 8 to the Posttest. An alpha level of 0.05 was used to conduct the paired t-tests

The overall mean difference in Reaction Time from Training Session 8 to Posttest is -0.99 seconds with a standard error of 0.27. About half the Participants did better on Training Session 8 than the Posttest. However, some of the who went very fast on Training Session 8 had a large increase in time on the Posttest. There is a significant difference between Training Session 8 and the Posttest, p-value = 0.0014.

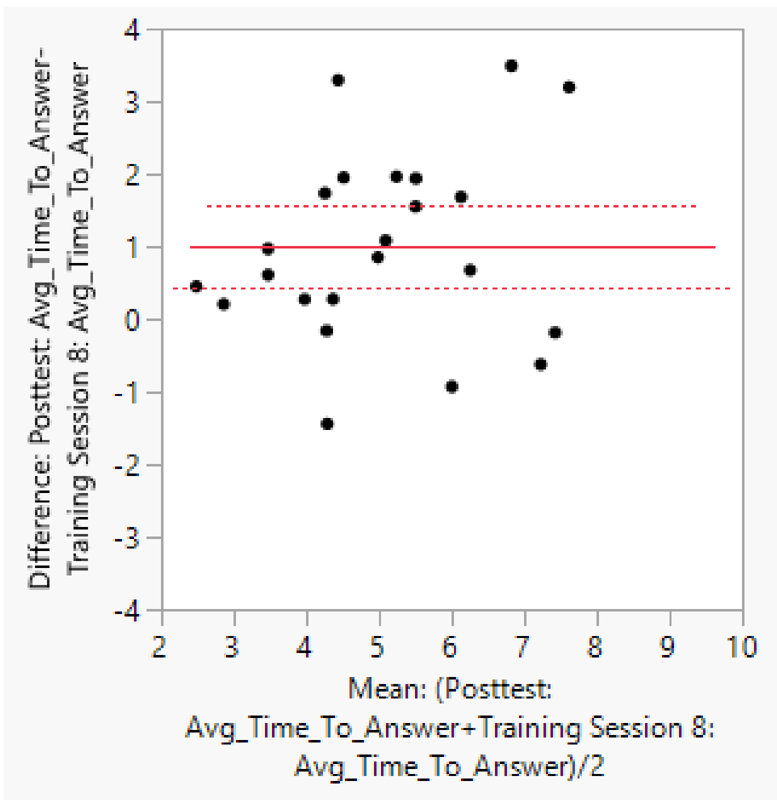


Figure 33. Matched Pairs: Difference in Mean Reaction Time from Training Session 8 to Posttest Graph

F. EXPLORATORY ANALYSES

1. Data Analyzed by Aircraft Category

Five categories of aircraft were tested, with eight aircraft in each category. To explore whether participants found certain categories more challenging than others, I conducted an analysis for the pretest and pretest – posttest score differences by aircraft category. The Scores for each Group by aircraft category are depicted in Figure 34. The difference in Score from Pretest to Posttest for each Group by aircraft category are depicted in Figure 35.

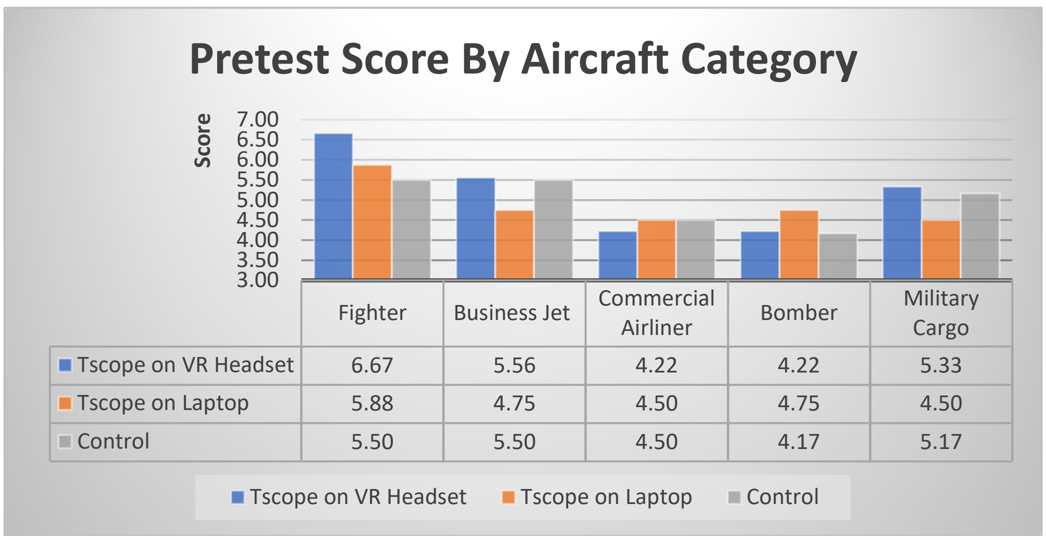


Figure 34. Pretest Score by Aircraft Category Graph

Regardless of group, it appears Participants had the least familiarity with Commercial Airliners and Bombers prior to training. Consequently, it is logical those categories showed the largest pretest - posttest differences. There is an inverse relationship between Pretest Score and Difference in Score, correlation of -0.86.

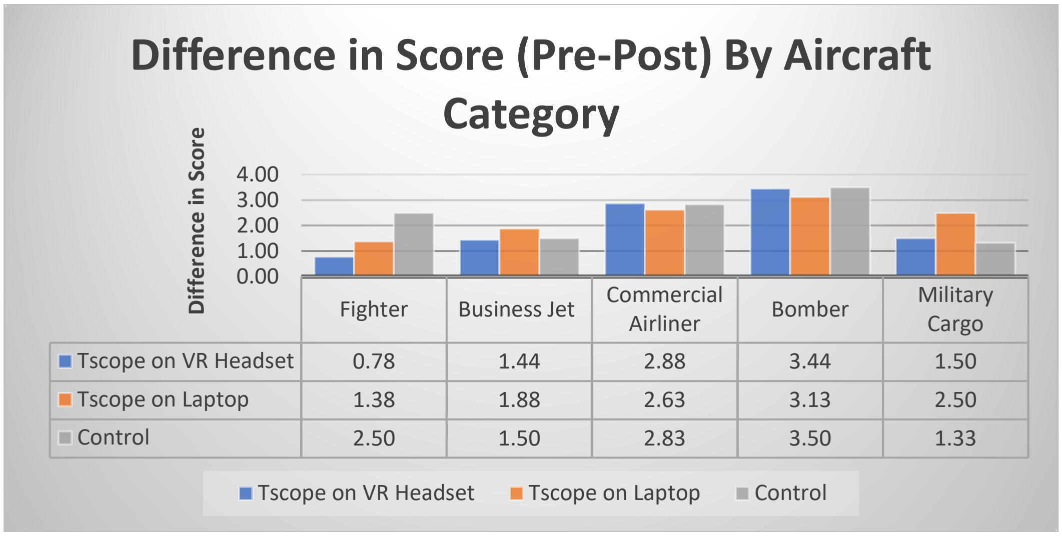


Figure 35. Difference in Score (Pre-Post) by Aircraft Category Graph

2. Pretest Score Prediction of Increase in Posttest Score, ANOVA Test

To explore whether those who performed more poorly on the pretest showed the largest increase on the posttest, a correlation was conducted to determine if there was a relationship between Pretest Score and the Difference in Pretest – Posttest Score. There is a correlation of -0.89 with a p-value < 0.0001, as seen in Figure 36.

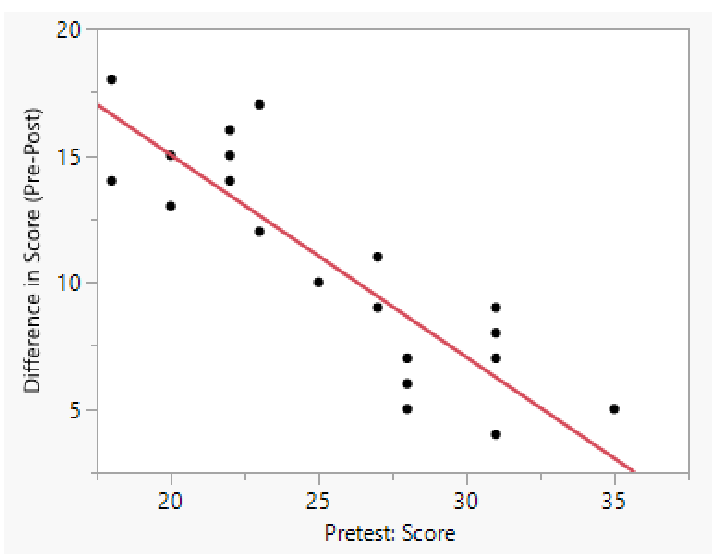


Figure 36. Bivariate Fit of Difference in Score (Pre – Post) By Pretest Score Graph

3. Confidence Level Overall Increase by Individual

To explore whether participants' confidence ratings changed over the course of the training, the researchers conducted a paired t-test on the difference in participants' mean pretest confidence level and mean posttest confidence rating. Results indicated that participants' confidence significantly improved ($t(22) = 11.41$, $p < 0.0001$). The Participants' confidence was low in the Pretest, Overall Average of 42.28 out of 100. The Overall Average Confidence Level on the Posttest, 85.79, was double that of their Pretest. Figure 37 illustrates changes in confidence levels across the three groups.

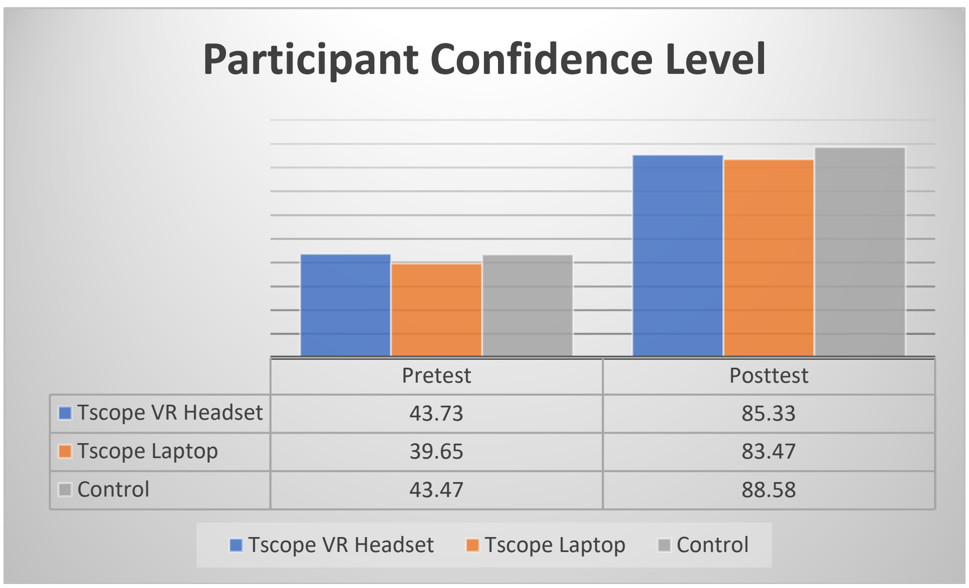


Figure 37. Participant Confidence Level Graph

4. Compare Results of Pretest to Training Session 8, Pair T-Test

To ascertain improvements in score and reaction time without the delay between Session 8 and the Post test, paired t-tests were conducted to determine if differences exist for Scores and Reaction Times from Pretest to Session 8.

a. Score

The difference in Scores, Pretest to Session 8, is somewhat normal and multimodal, as seen in Figure 38. The mean difference in Score is 13.1 from Pretest to Session 8 with a standard deviation of 5.0.

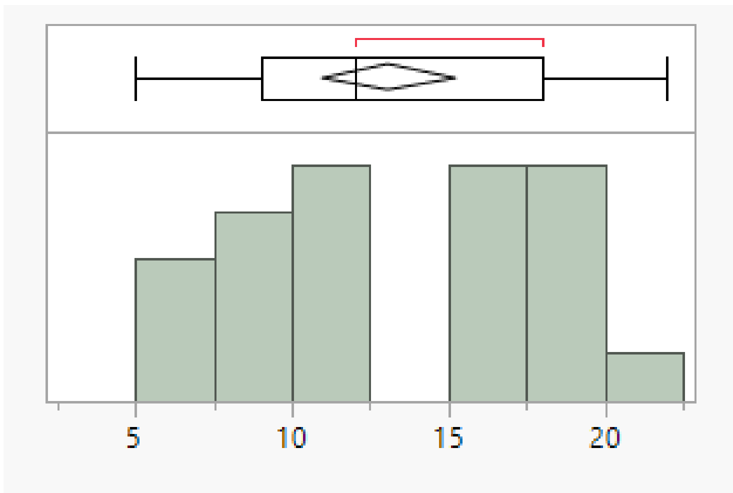


Figure 38. Difference in Mean Overall Score Pretest to Session 8 Graph

Table 9. Difference in Mean Score (Pre – Session 8) Summary Statistics

Mean	13.086957
Std Dev	4.9901088
Std Err Mean	1.0405096
Upper 95% Mean	15.244841
Lower 95% Mean	10.929072
N	23

Each Participant's Score increased from the Pretest to Session 8. The smallest increase was five. Three participants tied for the smallest increase; all those participants had Pretest scores over 30. The difference in Score from Pretest to Session 8 had a p-value < 0.0001, the difference is statistically significant, see Figure 39.

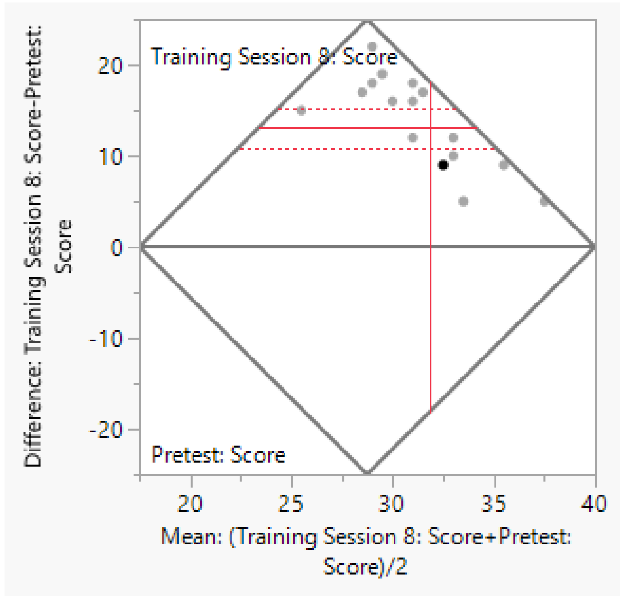


Figure 39. Matched Pairs: Pretest to Session 8: Score Graph

b. Reaction Time

The difference in Reaction Time, Pretest to Session 8, is mostly normal, as seen in Figure 40. There is one outlier. The participant was a foreign officer with limited knowledge of U.S. aircraft. The mean difference in Reaction Time is -1.30 seconds from Pretest to Session 8 with a standard deviation of 1.82.

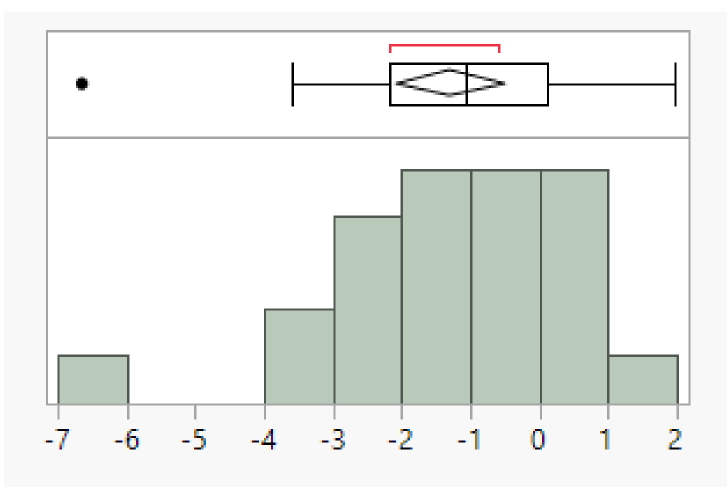


Figure 40. Difference in Reaction Time (Pretest – Session 8) Summary Statistics Graph

Table 10. Difference in Reaction Time (Pre – Session 8) Summary Statistics

Mean	-1.298629
Std Dev	1.8226099
Std Err Mean	0.3800404
Upper 95% Mean	-0.510474
Lower 95% Mean	-2.086785
N	23

All but six of the 23 participants lowered their Reaction Time from the Pretest to Session 8. Of note, the Tscope on VR Headset Group completed Session 8 on the VR Headset, not on the laptop, and the Reaction Time of that group was higher in every test and training session when compared to the other two groups. The difference in Reaction Time from Pretest to Session 8 had a p-value = 0.0025, the difference is statistically significant.

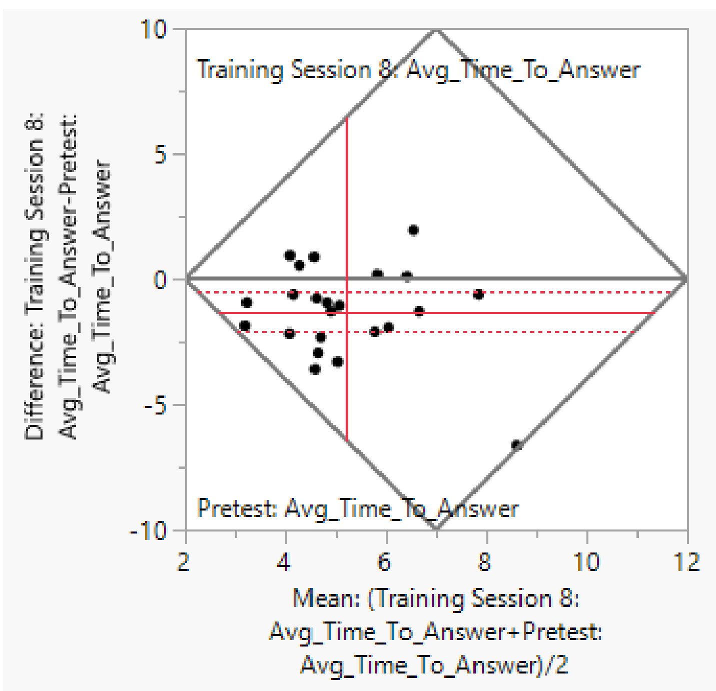


Figure 41. Matched Pairs: Pretest to Session 8: Reaction Time Graph

G. POST EXPERIMENT SURVEY ANALYSIS

1. Difficulty in Memorizing Aircraft

The Post Task Survey asked participants to rate how hard it was to memorize all 40 aircraft. The results are depicted in Figure 42. There was a mostly normal distribution around average difficulty. The Control Group reported an even spread from Very Easy to Moderately Difficult. Surprisingly, the Tscope on Laptop Group reported the task of memorization less difficult than the Tscope on the VR Headset Group.

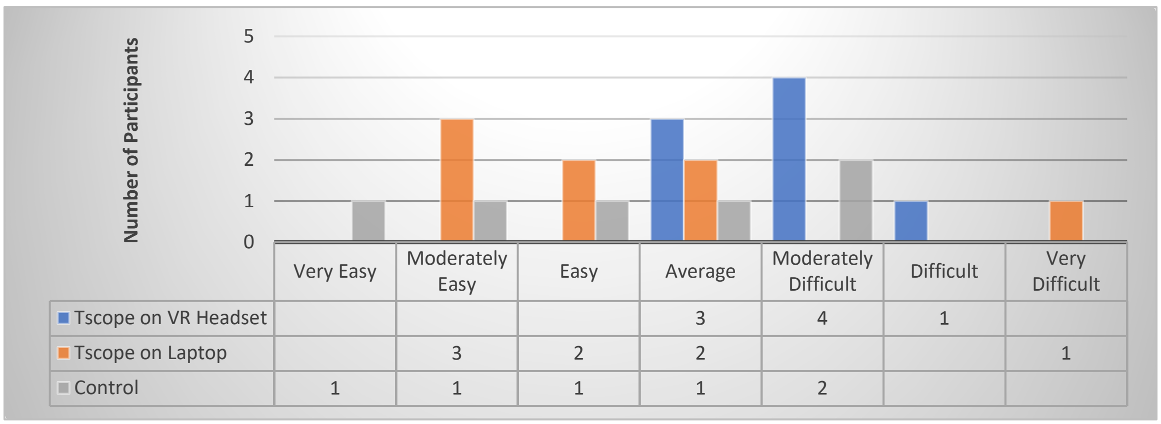


Figure 42. How Difficult Was it to Memorize Aircraft

2. Experimental Memory Method Compared to Previous Methods

The standard way in which most participants would study aircraft and other objects are flashcards, and power point. A few used mnemonic devices and word shape associations. One participant reported using the Army's ROC-V program. Participants were asked if they thought they could remember more aircraft using previous methods. The overall distribution was fairly normal and the distribution across groups was even. Further studies should use actual Power Point slides and flash cards in the Control Group to fully evaluate the potential benefits of the Retention Software Program.

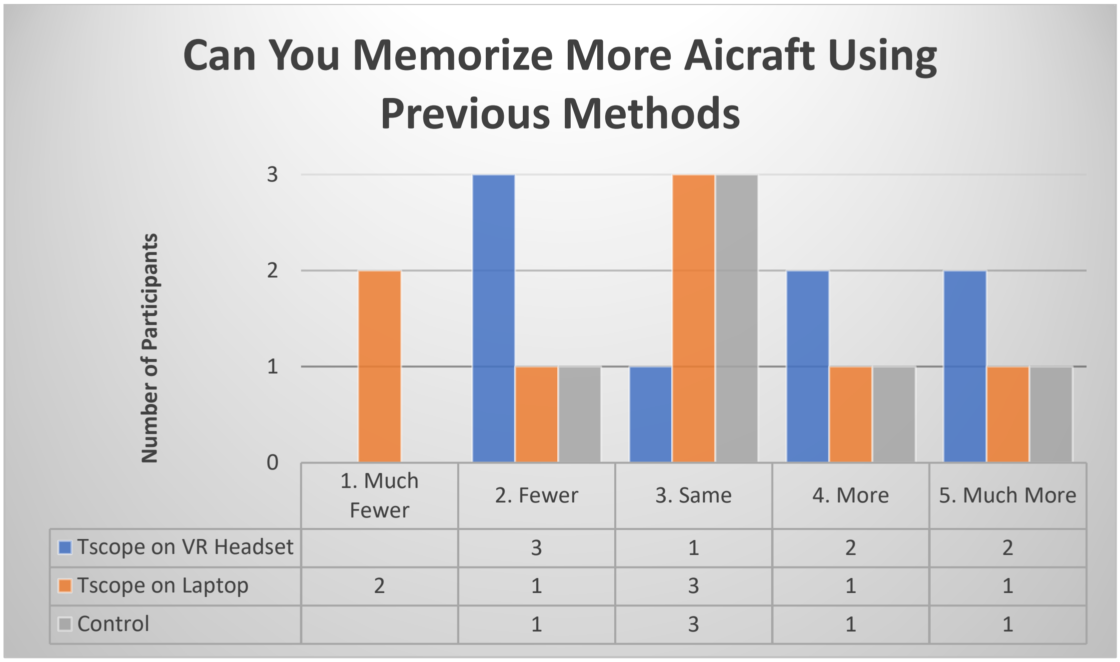


Figure 43. Can Participants Memorize the Same Number of Aircraft with New System Compared to Previous Methods

3. Parts of Aircraft Used to Identify Aircraft

Participants were asked to identify how they viewed and identified the aircraft. They were provided a list of options to choose from: Whole Shape of Aircraft, Wings, Engine, Fuselage, Tail, or Other. Only one participant chose one option, Whole of Aircraft. The other participants said they used multiple methods. Further studies using eye tracking would be required to confirm the participants answers.

H. GENERAL OBSERVATIONS

1. Guessing

Multiple Participants stated they guessed on a lot of Pretest questions, some of whom answered more than 30 correct. The low confidence levels on the Pretest support these statements.

2. Ceiling Effect on Training Session 8 and Posttest

The overall average for Training Session 8 was 38.3 out of 40. Eight out of 23 Participants received a perfect score on Training Session 8, suggesting a ceiling effect. If the study included more aircraft, it is likely the average number of correct answers would be even higher.

The Posttest had less of a ceiling effect. There were only three perfect scores. This decline from session 8 presumably is due to the four-day delay.

3. Issues with VR Headset

The VR Headsets began to crash mid-training on Training Session 6. The cache was not clearing and caused the memory to crash. This is still a bug in the current system. Once the program was reset, the cache would clear and start working properly.

4. General Comments from Participants

- Adaptive learning algorithms would make the training much better. Once a participant gained mastery of an aircraft, they did not like that the aircraft still received the same amount of training time.
- Participants liked the flashing, it helped them focus in on a particular part of the airplane.
- Multiple perspectives of the aircraft helped the participants remember better.
- Add a side-by-side comparison mode.

5. Physical Impacts of Tachistoscope

A majority of the tachistoscope participants stated they had no signs of nausea, eye strain, or other discomforts. While seven out of 16 tachistoscope participants agreed the flashing caused some level of eye strain or other discomfort; most all of them said the discomfort and eye strain decreased the more they conducted the training. Some said it would cause minor eye strain if the training was too long and only one participant said the VR headset caused light nausea.

I. SUMMARY

The Results Chapter provided a detailed account of the data collected throughout the study. Since there was no difference in Score or Reaction Time, neither Hypothesis 1 nor Hypothesis 2 could be rejected. Even so, all the Participants showed significant improvement in Score and Confidence Level.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. CONCLUSIONS

This chapter will cover a summary of the results, limitations, recommended areas of future work, and conclusions of key findings.

A. SUMMARY OF RESULTS

All participants showed significant improvement from Pretest to Posttest in Score, but not in Reaction Time. Participants with lower Pretest Scores showed even greater improvement than those with higher Pretest Scores. However, participants did show greater improvement from Pretest to Training Session 8 in both Score and Reaction Time.

The design of the experiment controlled for Participants' experience with aircraft, time in service, and previous vehicle recognition training. Across each of the groups, Participants' had similar demographics.

Participants gained significant confidence in their ability to accurately identify aircraft. On a scale of 0 to 100, the Average Confidence Level increased by 43.5 across all three groups. The Confidence Levels had high correlations for both the Pretest and the Posttest, 0.82 and 0.58 respectively. These results suggest the Participants were aware when they were guessing and when they knew the correct answers. Military members need to distinguish between when they know the answer and when they need to ask for assistance. This evidence supports the idea the Retention Software Program can help build accurate confidence as well as increase users' ability to correctly identify objects.

B. DISCUSSION

1. Time Limit on Study Material Is a Positive Factor

If the amount of time an object can be observed is restricted, the learners' attention is potentially more focused. The learner knows the object will only be present for a limited time; therefore, they must pay attention, or they will lose their ability to view the object. It is conjectured this was the effect seen in the control group. Even though the image was not flashed at a rapid pace, the participants were only able to view each image once for five seconds. In addition, once the participants clicked to start the object training, they did not have control of the timing.

2. Blank Interstimulus Interval

In this study, the blank ISI for the Tachistoscope Groups was 500 ms. Potter and Fox (2009) determined the ISI is an important component in an individual's ability to remember images. Their study showed memory performance increased from 20% correct with no ISI to 84% correct identification with an ISI of 1.5 seconds (Potter & Fox, 2009). This adjustment to the ISI could potentially further increase the learning rate of individuals in the Tachistoscope Groups.

3. Transfer Training

Multiple Participants made comments about identifying aircraft in other settings. One was eating an airport watching airplanes take-off and land. While watching the aircraft, he realized he was capable of distinguishing specific characteristics and identifying aircraft that were included in the training set. Another Participant stated they watched aircraft flying over-head and guessed the type. The Participant then checked their phone to confirm their guess and they were correct.

Transfer training was not tested during this experiment. All information on transfer training was self-reported by participants.

4. Participants Guessed on Many Pretest Questions

Multiple Participants stated they made a lot of guesses on the Pretest. The low Confidence Levels on the Pretest (42.3) support this claim. Only six Participants had a Confidence Level above 50, while only five Participants Scored at or below 50%. Either Participants were more capable than confident, or they guessed above average. The mean Score was 25.3 with a standard deviation of 5.0. The true 50% mark of the Pretest was 20, which is within 1.06 standard deviations of the mean.

5. Ceiling Effect

In the Posttest and all but three Training Sessions there were at least two perfect scores. Training Session 8 had the most perfect scores with eight. The number of perfect scores would reason those participants did not reach their learning capacity. Furthermore, the number of perfect scores increased even as the quantity of aircraft in the immediate tests increased. While most participants did not reach a perfect score when tested on all 40 aircraft, their scores in general showed progress. That progress indicates all the participants could be taught to recognize at least 40 aircraft with additional training.

6. Loss of Attention, Length of Training Session

When the training sessions become too long it can affect the participants ability to focus and maintain attention. The Pretest and the Posttest and first week training session took around 8 to 15 minutes. Participants began to complain around Training Session 6 which took about 20 – 25 minutes and had 35 aircraft. Some participants said their mind would wander and they would lose focus.

To alleviate this issue, it is recommended the training does no go over 20 minutes. The tachistoscope program could also add in more interactive features to keep the participants more engaged.

7. Reaction Time Could Have Decreased if Participants Knew It Was Being Measured

During the study, participants were not informed their Reaction Time was measured until after the study concluded. Participants made comments they took their time answering questions because they did not think Reaction Time was relevant. The Reaction Time difference from Session 8 to the Posttest (0.92 seconds) was the was only one of two increases; the other increase in Reaction Time was from Session 1 to Session 2 (1.79 seconds). There is a potential participants took longer on the Posttest because they thought the Posttest was more important than the training sessions and Score was the only metric being measured. Reaction Time was significantly quicker from the Pretest to Training Session 8, as discussed in paragraph V.F.4.b and seen in Figure 40; which supports this claim.

8. Sustainment Training

The dip in Scores and increase in Reaction Time on the Posttest was expected due to the delay between training and testing. Even still, the change was small. It is conjectured sustainment training would increase Scores and decrease Reaction Time. This is a potential topic to test in the future.

C. IMPLICATIONS

Results suggest that individuals can be trained to recognize objects in a relatively short period of time. With a proper user interface, the Retention Software Program can be adapted to suit any set of physical objects. Training time on the Retention Software Program can easily be tailored to meet the training requirements and time constraints of the user. This flexibility allows the training sessions to be easily scheduled into a unit's training plan or an individual's personal schedule. In addition, the Retention Software Program can be run on any modern computer and is therefore portable.

This system would be useful in initial schoolhouses, scout and reconnaissance units, tow gunners, pilots, navigators, forward observers, imagery analysts, or any specialty that requires someone to recognize an object or person. If a unit is deploying to a new Area of Operation, the commander and his staff could use the Retention Software Program to memorize important friendly and enemy personnel and equipment.

While the focus of this study is on aircraft in daylight conditions, it opens a path to pursue research on facial recognition as well. The Retention Software Program was created during this study and that same platform could be used to train commanders and their staffs to recognize friendly and enemy personnel prior to entering an area of operation. If on the day of arrival, a commander and his staff can engage coalition partners by name, it builds instant credibility. In addition, it frees up their cognitive load to focus on more complex tactical and operational requirements that demand their attention.

D. LIMITATIONS

When an individual identifies an aircraft in real life situations, there are variables they may encounter that were not captured in this study. For instance, the individual must recall the name of the aircraft because it would not be presented to them. Or a new aircraft is perceived that was not included in the training set. Moreover, the aircraft could be occluded by other objects that may change the appearance.

1. Visual Perspective

In addition, images of the aircraft often lack size perspective. While a C-17 and C-5 have similar shapes, they are drastically different in size. This is very apparent when you view them side-by-side in real life. It may be useful to add images with more perspective once a user can accurately identify the aircraft; so not to confuse the aircraft with other objects.

2. Hardware Requirements

To use the Retention Software Program, one needs either a VR headset or a computer. There will be times when personnel cannot readily access either one. In these circumstances they may need to revert to Flash Card or Power Point slides to study objects.

3. Image Database

Building the image database for the objects can be time consuming. All objects must be screened for accuracy, size, and quality. If multiple angles are required, it can be difficult to find adequate images.

E. FUTURE WORK

1. Images in Pretest and Posttest Are Quickly Presented

Future research should be conducted to evaluate a Participant's ability to rapidly recognize an object, regardless of platform. The Pretest and Posttest should rapidly present objects to limit the time a Participant is able to observe the object during the test. This would better evaluate a Participant's Reaction Time before and after training on the tachistoscope.

2. Control Group Using Flash Cards and Power Point

In a future study, the Control Group should use Flashcards and/or Power Point while the experimental groups use the tachistoscope. This will help evaluate if the tachistoscope is truly a better learning tool than two of the most used methods.

3. Update Tachistoscope

a. User Interface (UI)

In order the make the Tachistoscope program usable for the general population, a simple, intuitive UI must be created. The UI will need the following requirements:

- Create Objects with tags
- Name and upload images, tag those images to an object
- Adjust presentation timing
- Autogenerate JSON files to run the Tachistoscope program
- Create usernames and password, tag members to a group
- Change timing configurations for individuals and groups (presentation time, presentation quantity)
- Create which objects will be in the training sessions
- Generate tests
- Create delay between training sessions

b. Add in Adaptive Learning Algorithm

An adaptive learning algorithm was intentionally left out of this study. Because the experiment showed positive results, an adaptive learning algorithm should now be added. One of the most common comments from Participants was they did not want to train on aircraft which they had mastered. They saw this as a waste of time and reduction in focused attention. An adaptive learning algorithm would allow trainers to shorten the training time while increasing the number of objects to be memorized.

c. Security Concerns, Does Not Allow Computer to Go to Sleep

If the Retention Software Program is going to be used on a DOD networked computer, it must be evaluated to information assurance and cyber security requirements.

d. Oculus Crashes

There is a bug in the Tachistoscope program that causes the system to crash on the Oculus around the 900th image presentation. Every image that is presented in the training session is saved in the cache. The system crashes once the cache memory is overloaded. A modification needs to be made to clear the cache.

4. Test for Transfer Training

Multiple Participants have stated they can more accurately identify aircraft in real world settings. Some have even stated they view aircraft differently and can identify specific differences in aircraft they have not been trained on. While this information is positive, there is no scientific evidence to back the claims of the participants. To truly determine the accuracy of their statements, further testing will need to occur.

5. Memory Decay and Sustainment Training

From Training Session 8 to the Posttest there was significant memory decay. Future testing should be conducted to determine how much forgetting occurs at distinct intervals (e.g., 15-days, 30-days, and 45-days). Once that is determined, further testing should be done to determine how much sustainment training is required to maintain proficiency.

6. Test for Reaction Time, But Tell Participants

Participants were not told their Reaction Time was being measured on every question. Once told Reaction Time was being measured, a few participants said they would have answered more quickly.

F. CONCLUSION

Learning to memorize large sets of objects and rapidly recognize those objects is difficult. The preponderance of individuals in the military use flash cards or PowerPoint slides with images. The military does not currently have an effective and adaptive system that can easily be customized for individual unit or personal use.

Even with advancements in technology, humans still have a requirement to recognize objects. AI and advanced sensors may reduce the need for humans to be the ones who first identify objects on the battlefield. However, humans will always have a requirement to build mental models of objects to understand what the AI has identified. Humans are visual creatures; if one cannot recognize an aircraft from a car, how can they remember the capabilities of the aircraft and how it impacts the battlespace?

Military members are oversaturated with more requirements than time allows; to increase performance and readiness, effective and efficient training tools are required. This experiment was designed to address those issues. It gives users significant control over what they will train on and for how long. While the results between groups did not differentiate, Participants' performance increased as the training progressed.

Current recognition training methods focus primarily on how to look at an object (Wing-Engine-Fuselage-Tail process) or pair written object capabilities with an image (PowerPoint). There is not a widely used system with proven effectiveness. Military members across the services want a more effective and efficient method to memorize objects.

The Retention Software Program created for this experiment, effectively trained individuals on object recognition using aircraft as its medium. In about two-and-a-half training hours, participants reached an average of 96% accuracy (sd = 1.8), with eight perfect scores on Training Session 8. After a four-day delay, participants still achieved an average of 90% accuracy (sd = 2.3) on the Posttest. While there was not a significant change in Reaction Time from Pretest to Posttest, on average, the difference from Pretest to Training Session 8 was significantly quicker by 1.3 seconds.

In addition to accuracy and Reaction Time, each question included a confidence scale. Confidence ratings were an accurate predictor of the participants accuracy, which provided researchers with insight into the level of guessing done by the participants. Based on Pretest comments and Confidence Levels, it appeared participants guessed a significant portion of the time. After eight training sessions, all Posttest Confidence Levels and Scores showed significant improvement. Not only were Participants getting more answers correct, but they were also answering correct with accurate confidence.

This thesis reviewed current and past methods used to train service members on the memorization and rapid recognition of objects. The researchers wanted to combine an older proven method, the tachistoscope, with modern technology. The Retention Software Program effectively reached this goal. It is capable of operating on multiple platforms and is highly customizable. The architecture has been laid; with a few software updates, units will be able to create tailored training packages for their unit and mission. This training tool can increase capability across a spectrum of units throughout the Department of Defense.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. TACHISTOSCOPE SPECIFICATION DOCUMENT

Needs Statement: Create a tachistoscope program that can operate on both a computer and in a VR headset. The user interface to load and tag images needs to be on the computer and then loaded onto the VR headset.

Tachistoscope Requirements:

1. The time an image is presented needs to be a parameter that can be adjusted through the computer interface.
2. The images need to be shown for a precise amount of time, down to 100ms.
3. The time the images are shown need to be timed and recorded.
4. Each user needs an individual login and profile
5. A log needs to be kept for each user that can be exported as a csv or excel doc
    - 5.1. Assign Learner unique ID
    - 5.2. Time user logged in to train
    - 5.3. Training conducted when user logged in
    - 5.4. Track score for game play
        - 5.4.1. Log reaction time when object was selected, from time image was presented to time answered
        - 5.4.2. Log how many right and wrong answers were selected
        - 5.4.3. Log score and save score
    - 5.5. Track incomplete training sessions
    - 5.6. Add confidence rating to how well the subjects believe they did on each question
6. If training session is incomplete, training for that session will be restarted
7. Add gamification to gameplay to keep subjects engaged in training
8. Database to add objects and images
    - 8.1. Database will be accessed through the computer interface
    - 8.2. Every object can have images assigned to themselves
    - 8.3. Drag and drop images to the object
    - 8.4. Objects and images will have tags
        - 8.4.1. Generic tags
            - 8.4.1.1. Day
            - 8.4.1.2. Night
            - 8.4.1.3. Ground vehicle
            - 8.4.1.4. Aircraft
            - 8.4.1.5. Ground Equipment
            - 8.4.1.6. Satellite
            - 8.4.1.7. Language
            - 8.4.1.8. Person
        - 8.4.2. Allow user to add custom tags
9. Tachistoscope Operation

9.1. Overall operation, see Figure 44
9.2. In computer interface, be able to create customizable training sessions
  9.2.1. For each training session, be able to assign objects to those sessions
  9.2.2. Select how many objects will be selected for that training session
  9.2.3. For each training session be able to adjust image presentation timing
    9.2.3.1. One setting for previously viewed objects
    9.2.3.2. One setting for new objects
  9.2.4. Option to randomly pick which object to train, or create order in which objects will be trained

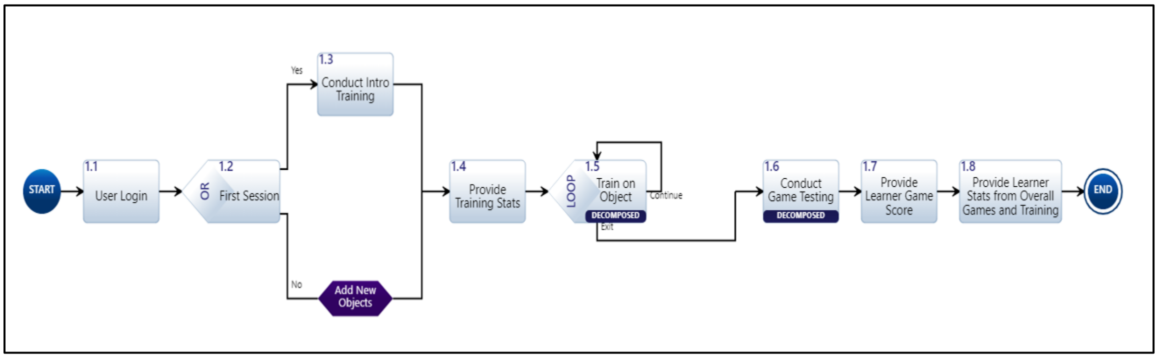


Figure 44. Tachistoscope Operation Diagram

9.3. Training Sessions: See Figure 45.

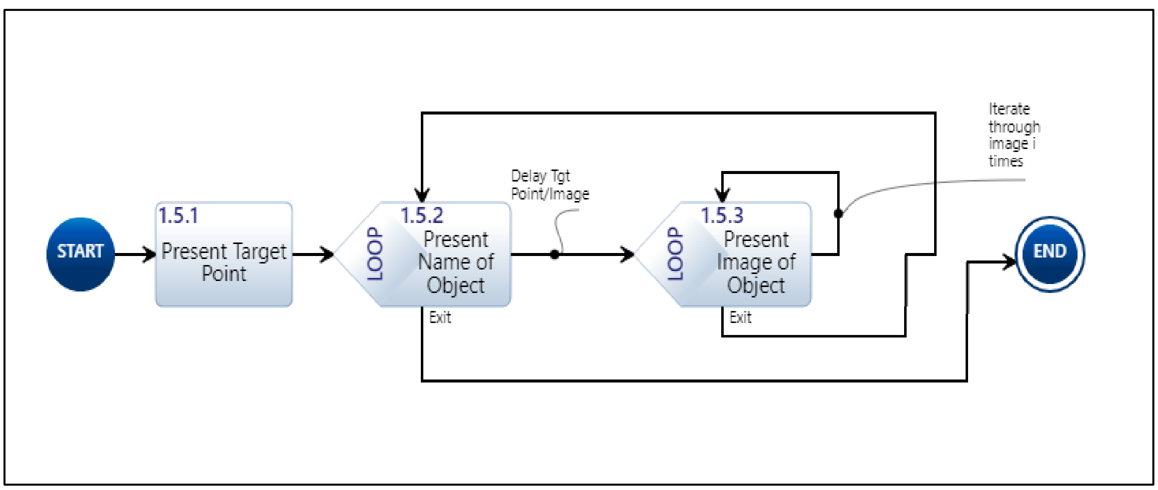


Figure 45. Testing Session Diagram

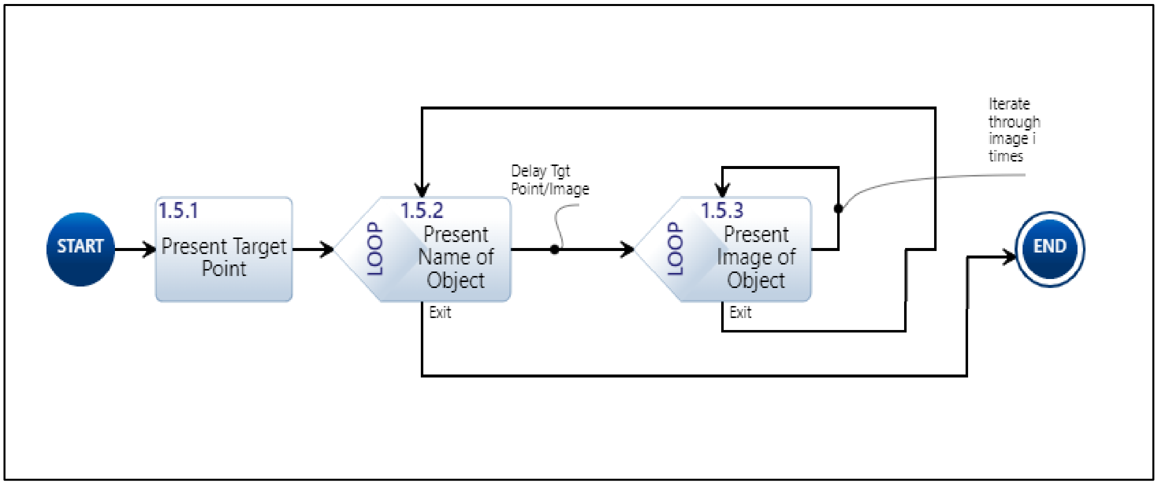


Figure 46. Testing Session

9.3.1. Training Session Parameters:
- 9.3.1.1. “targetPointDisplayTime”: 1.5.1: Default display 1000ms
- 9.3.1.2. “objectNameDisplayTime”: 1.5.2: Default display 2000ms
- 9.3.1.3. “imageDisplayTime”: 1.5.3 Display Time for Image: Default 250 ms
- 9.3.1.4. “imageHideTime”: Delay between 1.5.2/1.5.3 and between iterations of 1.5.3: Default 500ms
- 9.3.1.5. “imageDisplayCount”: 1.5.3 Iterations Image is Displayed: Default 5

9.4. Game Play

- 9.4.1. For computer interface, setup how many iterations of game play will occur in each session
- 9.4.2. Name of object and one image will be presented
- 9.4.3. Learner will hit a button for “Yes” if Object and Image match, another button for “No” if they do not match
- 9.4.4. Time from image presentation to button push will be saved in log, Reaction Time
- 9.4.5. Correct and incorrect answers will be saved in log
- 9.4.6. Matching Images will be shown in random order
- 9.4.7. Non-Matching images will be selected randomly from pool of same tags or category
- 9.4.8. Sequence for Game Play is shown in Figure 47 Game Play Diagram.

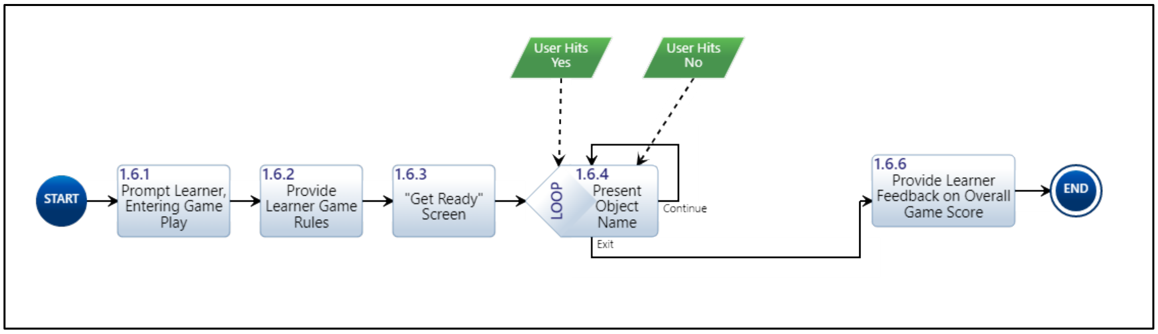


Figure 47. Game Play Diagram

10. Learner Feedback. At the end of each session, the learner will be provided feedback
    - 10.1. Training session completed
    - 10.2. Number of objects the learner correctly identify
11. Pre-Test/Post-Test
    - 11.1. Same structure as the Game Play, Figure 3.
    - 11.2. Separate Module that will allow researcher to use separate pictures from the training
    - 11.3. Log results of Pre-Test and Post-Test of each user
    - 11.4. Put admin controls on Pre-Test and Post-Test to prevent subjects from accessing during training

APPENDIX B. SAMPLE TRAINING LOG OUTPUT

At the conclusion of each training session and testing session a csv file was automatically generated and forwarded to the Amazon Web Services (AWS) cloud.

Table 11. Sample Training Log Output

0.009246	start-time	12:20:47
0.013022	User Logged In	User100
0.042827	Tutorial Started	
4.812246	Tutorial Step Completed	1
8.580617	Tutorial Step Completed	2
11.113966	Tutorial Step Completed	3
15.54136	Tutorial Step Completed	4
21.026934	Tutorial Step Completed	5
21.034824	Tutorial Completed	
24.466365	Training Session Started	Training Session 1
24.466754	Training Started	
386.251043	Training Completed	
388.642079	Test Started	C-17
393.83112	Test Submitted Answer	No
393.83149	Test Graded	Correct
393.833651	Time On Test (in seconds)	5.189384
399.014009	User Confidence Level (0-100)	98.08
399.014472	Test Started	C-2
415.672422	Test Submitted Answer	No
415.672459	Test Graded	Correct
415.672516	Time On Test (in seconds)	16.65799
420.059159	User Confidence Level (0-100)	50
420.059429	Test Started	CRJ1000
427.733907	Test Submitted Answer	No
427.733945	Test Graded	Incorrect
427.733986	Time On Test (in seconds)	7.674509
432.56949	User Confidence Level (0-100)	71.35
432.569739	Test Started	787
436.574349	Test Submitted Answer	Yes
436.574382	Test Graded	Correct
436.574434	Time On Test (in seconds)	4.004643
439.747022	User Confidence Level (0-100)	71.22

0.009246	start-time	12:20:47
439.747285	Test Started	BAE-146
443.107362	Test Submitted Answer	No
443.107398	Test Graded	Correct
443.107444	Time On Test (in seconds)	3.360099
448.880478	User Confidence Level (0-100)	86.06
448.880735	Test Started	DF900
456.089297	Test Submitted Answer	Yes
456.089335	Test Graded	Correct
456.089377	Time On Test (in seconds)	7.208594
461.319341	User Confidence Level (0-100)	29.06
461.319636	Test Started	Mirage-2000
465.772374	Test Submitted Answer	Yes
465.772413	Test Graded	Correct
465.772459	Time On Test (in seconds)	4.452777
470.32405	User Confidence Level (0-100)	74.51
470.324315	Test Started	F-35
473.748761	Test Submitted Answer	Yes
473.748798	Test Graded	Correct
473.748854	Time On Test (in seconds)	3.424479
477.572714	User Confidence Level (0-100)	100
477.572965	Test Started	B-2
479.942954	Test Submitted Answer	Yes
479.942991	Test Graded	Correct
479.943042	Time On Test (in seconds)	2.370025
484.286521	User Confidence Level (0-100)	96.26
484.286783	Test Started	Tu-160
491.884528	Test Submitted Answer	Yes
491.884566	Test Graded	Correct
491.884623	Time On Test (in seconds)	7.597786
494.27381	User Confidence Level (0-100)	29.4
497.545405	Training Session Completed	Training Session 1
497.548319	end-time	12:29:05

LIST OF REFERENCES

Admin. (2019, December 18). *Tachistoscope Method—Photographic Memory Training Using Flash Cards*. Rightbraineducation. https://www.rightbraineducationlibrary.com/post/tachistoscope-principle

Antonov 225—Russian Transport Aircraft. (2007, August). Defence Forum & Military Photos - DefenceTalk. https://www.defencetalk.com/military/photos/antonov-225-russian-transport-aircraft.25145/

Axel, J. (n.d.). *Hawker Beechcraft Corp Beechcraft B200GT King Air*. Retrieved March 25, 2022, from https://aerocorner.com/aircraft/beechcraft-king-air-b200gt/

Bauer, B. (2015). *A Timely Reminder About Stimulus Display Times and Other Presentation Parameters on CRTs and Newer Technologies.* Canadian Journal of Experimental Psychology, 69(3), 264–273.

Cave, C. B. (1997). *Very Long-Lasting Priming in Picture Naming.* Psychological Science, 8(4), 322–325. https://doi.org/10.1111/j.1467-9280.1997.tb00446.x

Donovan, J. J., & Radosevich, D. J. (1999). *A meta-analytic review of the distribution of practice effect: Now you see it, now you don't.* Journal of Applied Psychology, 84(5), 795–805. http://dx.doi.org/10.1037/0021-9010.84.5.795

Elgendi, M., Kumar, P., Barbic, S., Newton, H., Abbott, D., & Cichocki, A. (2018). *Subliminal Priming—State of the Art and Future Perspectives.* Behavioral Sciences, 8(6), 54. http://dx.doi.org.libproxy.nps.edu/10.3390/bs8060054

Fleaglass. (n.d.). Fleaglass. Retrieved September 3, 2021, from https://www.fleaglass.com/ads/ca-1880-smith-college-lab-collection-electro-mechanical-pendulum-possibly-for-early-sublimation-experiments/

Gibson, J. J. (1947). *Motion Picture Testing and Research*. U.S. Government Printing Office.

J-20—Fighter aircraft in flight. (2011, August 30). Defence Forum & Military Photos - DefenceTalk. https://www.defencetalk.com/military/photos/j-20-fighter-aircraft-in-flight.45959/

Kennedy, Q., Taylor, J., Noda, A., Yesavage, J., & Lazzeroni, L. (2015). *The STEP Model: Characterizing Simultaneous Time Effects on Practice for Flight Simulator Performance Among Middle-Aged and Older Pilots.* Psychology and Aging, 30. https://doi.org/10.1037/pag0000043

Kim, J. W., Ritter, F., & Koubek, R. (2013). *An integrated theory for improved skill acquisition and retention in the three stages of learning.* Theoretical Issues in Ergonomics Science, 14. https://doi.org/10.1080/1464536X.2011.573008

Klein, G. (2008). *Naturalistic Decision Making*. Human Factors, 50(3), 456–460. https://doi.org/10.1518/001872008X288385

Klein, G. A., & Calderwood, R. (1996). *Investigations of Naturalistic Decision Making and the Recognition-Primed Decision Model.* KLEIN ASSOCIATES INC YELLOW SPRINGS OH. https://apps.dtic.mil/sti/citations/ADA310303

Milgram, P. (1986). *A spectacle-mounted liquid-crystal tachistoscope.* Behavior Research Methods, Instruments, & Computers, 19(5), 449–456.

Oculus Device Specifications | Oculus Developers. (n.d.). Retrieved April 5, 2022, from https://developer.oculus.com/resources/oculus-device-specs/

Patterson, R. E., Pierce, B. J., Bell, H. H., & Klein, G. (2010). *Implicit Learning, Tacit Knowledge, Expertise Development, and Naturalistic Decision Making.* Journal of Cognitive Engineering and Decision Making, 4(4), 289–303. https://doi.org/10.1177/155534341000400403

Potter, M. C., & Fox, L. F. (2009). *Detecting and remembering simultaneous pictures in a rapid serial visual presentation.* Journal of Experimental Psychology: Human Perception and Performance, 35(1), 28–38. https://doi.org/10.1037/a0013624

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). *Detecting meaning in RSVP at 13 ms per picture.* Attention, Perception & Psychophysics, 76(2), 270–279. https://doi.org/10.3758/s13414-013-0605-z

Quick, Total, Accurate Recognition. (1943). Newsletter - Bureau of Aeronautics Navy Department, 185, 8–9.

Renshaw, S. (1945). *The visual perception and reproduction of forms by tachistoscopic methods.* Journal of Psychology, 20, 217–232.

Rowan, C. P. (2015). *Assessing memory decay rate: What factors are the best predictors of Decrements in Training Proficiency in a Threat Vehicle Identification Task?* http://journals.sagepub.com/doi/10.1177/1541931215591391

Science: Fast Looks. (1945). TIME; http://content.time.com/time/magazine/article/0,9171,886583,00.html. http://content.time.com/time/magazine/article/0,9171,886583,00.html

Sperdin, H. F., Repnow, M., Herzog, M. H., & Landis, T. (2013). *An LCD Tachistoscope with Submillisecond Precision.* Behavior Research Methods, 45, 1347–1357. https://doi.org/10.3758/s13428-012-0311-0

Tc3_20x31_1.pdf. (n.d.). Retrieved September 3, 2021, from https://armypubs.army.mil/epubs/DR_pubs/DR_c/pdf/web/tc3_20x31_1.pdf

Turner, L. (2022, July 28). *F-15 Eagle* [Military]. Air Force. https://www.af.mil/About-Us/Fact-Sheets/Display/Article/104501/f-15-eagle/

Vicory, A. C. (1968). *A Brief History of Aircraft Identification Training*. George Washington Univ Alexandria Va Human Resources Research Office. https://apps.dtic.mil/sti/citations/AD0676791

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California