Beyond the Cockpit: VR Training Solutions for Air Refueling Proficiency

Lindsay Gouedy	Mary Fendley, Ph.D.	Brandon "Steppen" Wolf, Lt
LTRI/Louisiana Tech University	Louisiana Tech University	Col, USAF
Ruston, LA	Ruston, LA	93rd Bomb Squadron
Lgouedy@latech.edu	mfendley@latech.edu	Barksdale AFB, LA
		Brandon.wolf.6@us.af.mil

ABSTRACT

Training military B-52 pilots is resource-intensive and time-consuming, necessitating extensive flight hours to master the skills for mission success. This research builds upon previous efforts aimed at validating Virtual Reality (VR) training for United States Air Force (USAF) pilots, particularly focusing on the air refueling (AR) task for the B-52 aircraft. The primary objective of this project is to evaluate whether USAF pilots trained using VR technology experience a reduction in training duration and/or the number of sorties needed to achieve proficiency compared to historical data from non-VR-trained pilots undertaking the same air refueling task. Data was analyzed from VR-trained against non-VR-trained B-52 pilots for time (in minutes) and sortie counts to proficiency. Eye-tracking technology and the NASA-TLX questionnaire were utilized to evaluate the cognitive load of the trainees across the four-step process between the VR and air environments. Differences in participants' cognitive load at different points during the training program and in different environments were observed enabling a deeper assessment of the environmental training impacts. Pre- and post-training questionnaires were employed to gather background information and subjective feedback from the trainees. A reduction in class averages of the number of sorties required to reach proficiency was observed with VR-trained classes.

ABOUT THE AUTHORS

Lindsay Gouedy is pursuing her Ph.D. in Engineering Education, serving as a National Security Fellow at Air Force Global Strike Command, specializing in virtual reality (VR) training methodologies. Concurrently, she works as a Doctoral Researcher and Adjunct Professor at Louisiana Tech University's DECODED Lab, focusing on VR flight simulation programs. She has also worked as Executive Director leading public relations and scientific research for the SGWCD for fifteen years. Lindsay holds a Master of Science in Engineering and Technology Management, a Master of Business Administration, and a Bachelor of Science in Agricultural Business. In addition, she holds black and green belt six-sigma certifications.

Dr. Mary Fendley holds degrees from Indiana University and Wright State University. She is currently an Associate Professor in Industrial Engineering at Louisiana Tech University. Her research interests are in the areas of cognitive engineering, modeling, decision analysis, training, trust in automation and human factors engineering. She is currently an Associate Editor of the Journal of Human Factors and Ergonomics in Manufacturing & Service Industries and serves as reviewer for numerous journals.

Brandon Wolf, Lt Col. is a Lieutenant Colonel in the U.S. Air Force Reserves, Director of Plans and Programs at Barksdale AFB, LA and an instructor/evaluator pilot in the B-52H Formal Training Unit. He has been the lead for the reserve wing's innovation efforts for four years and continues to develop and integrate new technologies and processes. He holds a Master of Military Operational Art and Science from Air University and a Bachelor of Science in Systems Engineering, Human Systems from the United States Air Force Academy.

Beyond the Cockpit: VR Training Solutions for Air Refueling Proficiency

Lindsay Gouedy	Mary Fendley, Ph.D.	Brandon "Steppen" Wolf, Lt
LTRI/Louisiana Tech University	Louisiana Tech University	Col, USAF
Ruston, LA	Ruston, LA	93rd Bomb Squadron
Lgouedy@latech.edu	mfendley@latech.edu	Barksdale AFB, LA
		Brandon.wolf.6@us.af.mi

INTRODUCTION

Many industries are keen on adapting to and incorporating tested technologies that prove to enhance their mission. The world of aviation is no different. The world of training simulation has evolved radically over the last century of flight training. From the invention and implementation of the first flight in the 1930s, known as the 'Link Trainer', to today's computer-based simulators, the objective is still the same, to provide pilots with the knowledge needed to operate an aircraft in a controlled environment (The Link Trainer Flight Simulator, 2020).

The training process of Air Force B-52 pilots has been governed by a syllabus that traditionally incorporates a mixture of classroom instruction, applied in-air practice flights (known as a 'sortie'), and practice in a WST (Weapons System Trainer) training simulator (a life-size motion simulator). Now, with the advancements in virtual reality (VR) technology used in the aviation world, as well as other defense areas, there has been a push to test whether this technology would benefit the training outcomes of the B-52 pilots. The research objectives for this project look to the next level of technological enhancements that will specifically improve air refueling (AR) training of B-52 pilots. First, we are attempting to gauge the efficacy of the use of VR training environments for air refueling, and secondly, identify and measure the differences in the cognitive load of pilots in VR training environments vs in-air training environments.

BACKGROUND

Air Refueling (AR)

Air refueling, also known as in-flight refueling or aerial refueling, is a critical military and aviation operation that involves transferring fuel from one aircraft (the tanker) to another (the receiver) while both are in flight (National Aviation Academy, 2023). This capability allows military and other aircraft to extend their range, endurance, and operational capabilities of military aircraft. It allows these aircraft to operate for longer durations without needing to land for refueling (National Aviation Academy, 2023).

Tanker aircraft are specially designed or modified aircraft equipped with fuel transfer systems. Today's common tanker aircraft include the Boeing KC-135 Stratotanker, Boeing KC-46 Pegasus, and the Boeing KC-10 Extender in the United States, as well as various models of the Airbus A330 MRTT (Multi Role Tanker Transport) in other countries (Airforce Technology, 2013). The KC-135, a tanker commonly used by U.S. Air Force, has a fuel-carrying capacity of over 30,000 gallons of fuel (Gentry, 1994). Receiver aircraft are typically fighter jets, reconnaissance planes, bombers, and other military aircraft. Some receiver aircraft are equipped with a refueling probe or receptacle to connect to the tanker's refueling boom or drogue.

There are two primary methods of air refueling: boom and drogue/probe refueling systems. In the boom refueling method, the tanker aircraft extends a rigid boom to make contact with a receptacle on the receiving aircraft (National Aviation Academy, 2023). Fuel is then transferred through the boom under the control of the boom operator. In the drogue/probe refueling method, a flexible hose with a drogue (basket) at the end is deployed from the tanker aircraft (National Aviation Academy, 2023). The receiving aircraft extends a probe to connect with the drogue, and fuel is transferred through the hose.

Air refueling for the Boeing B-52 Stratofortress is a critical capability that allows this long-range strategic bomber to extend its mission endurance and operational reach. Assigned operationally to Barksdale AFB and Minot AFB, the B-52 has been in service since the 1950s. The B-52 refuels using the boom refueling method and is most commonly refueled by a KC-135 Stratotanker or KC-10 Extender. The B-52 features a receptacle for aerial refueling located on the upper part of its fuselage, typically near the rear of the aircraft, and a fuel-carrying capacity of approximately 50,000 gallons (Davis, 2024). The tanker's aircraft boom operator must carefully align the boom with the B-52's receptacle to establish a fuel transfer connection.

The process of air refueling for the B-52 involves precise coordination between the tanker's crew and the bombers. This technology's capability has revolutionized military operations over the past century, but ultimately, it's the men and women in the aircraft that make this task happen (Gentry, 1994). So much of the success of this task by the operators within the aircraft relies on visual cues. This creates a challenging training situation for airmen in an air environment.

Virtual Reality Learning Environments

Historically, pilot training in the U.S. Air Force has been a mixture of academic and applied skill training. As technology has evolved, training methodology has remained largely conventional in format and application. However, there has been a revolving initiative to implement and determine the effectiveness of VR when training USAF pilots, starting with Pilot Training Next (PTN). PTN's focus was predominately on understanding the learning process of Airmen, and 'prototyping training environments' incorporated with different technologies to 'accelerate' the pilots' learning process (Hawkins, 2019). Since the inception of the program, there have been multiple iterations of the VR-based training learning projects, now called Project Fusion (Hunter, 2023). While the use of enhanced training technology is gathering more steam, researchers are looking deeper into the physiology of VR as a learning and training tool to test the efficacy of its implementation. Having a research-based understanding of both psychology and technology should provide a positive experience (Budgeon et al., 2022). In 2021, Van Weelden's (2021) research examined the emerging field of VR for use in aviation training ultimately resolving that this tool offers the potential to enhance skill development. A positive impact on pilot cognition is apparent as a research-based understanding of both psychological and technological improvements (Budgeon et al., 2022).

Interest in VR technology isn't limited to defense aviation, as a broad spectrum of commercial aviation organizations are incorporating more and more aspects of VR. In 2021, the European Union Aviation Safety Agency (EASA) gave the first regulatory approval for the use of VR flight simulation technology as an approved training device in Europe (Firsing, 2024). Two years later, in 2023, the FAA began evaluating the same VR simulation device from Loft Dynamics for regulatory approval in the U.S.(Loft Dynamics, 2023). The defense aviation industry is continuing to embrace the technological advancement that VR simulation provides. The U.S. Military budgets an estimated $14 Billion on virtual training programs (Garcia & Winer, 2022)). With the implementation of the Pilot Training Next (PTN), then Pilot Training Transformation (PTT), and now Project Fusion, the Air Force is actively testing the benefits of incorporating more VR training programs as a way to speed up pilot training and improve cost-effectiveness (Hawkins, 2019; Hunter, 2023). The focus of PTN was predominantly on understanding the learning process of Airmen, and "prototyping training environments" incorporated with different technologies to accelerate the pilots' learning process (Hawkins, 2019). The ability to manipulate the training environment to any scenario in a controlled environment is a clear advantage to incorporating VR in defense training (Garcia & Winer, 2022). Van Weelden also references the benefits of safe and controlled environments as a result of VR training (Van Weelden et al., 2021).

While VR use is seeing successful implementation for flight training, there is still a gap in specific flight tasks that this equipment could fill. Within the Air Force there are different areas of specialized flight training and with it unique tasks that are not covered in the broad flight training programs. Air refueling is an example of one such task which may be suited for the use of VR. There are currently minimal ground-based training applications for the very complex and dynamic task of air refueling.

Cognitive Load

The American Psychological Association defines cognitive load as the mental demands imposed on a person while performing a task. These demands increase significantly when the task is challenging, dangerous, or unfamiliar, leading to situations where performance speed and success can be greatly impacted (American Psychological

Association, 2018). Air refueling is a highly cognitively demanding task requiring precise coordination, real-time decision-making, and exceptional spatial awareness, all under the pressure of maintaining aircraft stability and safety. Pilots must manage multiple variables concurrently, such as monitoring fuel transfer rates, maintaining precise positioning relative to the tanker aircraft, and communicating with both the tanker crew and their own flight team. These activities demand strong concentration and rapid information processing, which can overwhelm even experienced pilots, particularly in poor weather conditions or during night operations. Specifically in the B-52, the cognitive demand is increased by the unique aspects of controlling a very large aircraft in a relatively small window behind the tanker. There are unique interrelated aerodynamic forces because of the large wingspan and significant delayed effects of the flight controls forcing the pilot to anticipate what control inputs to make and ensuring those inputs are not left in too long. Air refueling in the B-52 is an especially difficult task even for experienced pilots.

The complexity of air refueling tasks stresses the importance of understanding cognitive load to improve training methods and enhance pilot performance. By examining the cognitive demands of these tasks, we can develop more efficient and effective training programs that not only model the physical aspects of refueling but also address the mental challenges pilots face. This approach can help in creating strategies to mitigate cognitive overload, such as breaking down the task into manageable components, using advanced simulators for realistic training, and incorporating stress management procedures.

Ultimately, by recognizing and addressing the cognitive load involved in air refueling, we aim to improve pilot proficiency, reduce errors, and increase overall mission success. This research highlights the critical need for comprehensive training programs that are conscientious of cognitive load with an equal emphasis on the physical demands on the pilot's performing air refueling operations.

Training Transfer Modeling

As we look to determine the potential validity and strength of technology training programs and if they improve the transfer of training, we must also take a deep dive into training methodologies used in technology-based learning environments. To this point, the implementation of a VR format into conventional military learning environments has been largely by trial and error and with little cohesion amongst the programs. With an increased focus now on effectively implementing this technology, the application of an instructional design framework will help achieve this goal. The overall objective of any training program is training transfer in order to create a more skilled operator. This is not limited to any specific task but rather consistent across all industries where education and training are concerned. Data tells us approximately 20% of trained skills transfer to persistent memory (Tonhäuser & Büker, 2016). Three different levels of training transfer inform how successful a training approach can be: organizational, learning field, and individual. All three levels are equally important which necessitates finding a model to support that mentality. Given this context of training transfer, especially as it relates to an applied task, it would be evident that a training model incorporating VR environments should be geared less towards the abstract, as is the norm in conventional training, and more toward a hands-on and authentic environment.

Lave's (1988) Situated Learning Theory (SLT) is one of the first models reviewed as an option that could support a VR training program as it identifies three levels of training transfer, all be it in a different framework: activities, identities, and relationships. These three levels underscore the emphasis on community and learning as being interconnected. This learning theory specifically defines that for successful training transfer to take place, it must be trained organically, from environmental and situational authenticity (Learning Theories, 2022). A 2010 study by Zheng identified that learners' task performance was distinctly correlated with situational learning, underscoring the differences in traditional and nontraditional learning environments and providing statistical support for this model concept. Another model examined was the Theoretical Model for Multicapable Airmen. It establishes SLT as a part of its foundation because it was developed specifically with immersive military training in mind. It was presented at the 2022 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) as an integrated approach that combines the SLT and Experiential Learning Theory (ELT). ELT denotes learning as being a perpetual process that evolves and grows as more experience is gained (Kolb, 1984). Unlike SLT, ELT speaks to the need for reflection and abstraction as being just as important as hands-on experience. The integrated concept was designed as a viable framework by which the Air Force could use to train Airmen specifically using immersive technology (Ayers & Hamilton, 2022). This theoretical model brings the need for a community-centered, experience-based, and abstract reflective training methodology all into one learning concept (Ayers & Hamilton, 2022). One of the greater points

made alongside this concept is the need for a consistent training framework where immersive technology is being incorporated in aviation.

The Four Component Instruction Design (4C/ID) is another training model that has been tested in a vocational skills training program that incorporated VR training methodology (Mulders, 2022). The 4C/ID model is made up of four key components with subcomponents to construct the complete model. The four key components are as follows; 1) design the learning tasks, 2) design the supportive information, 3) design procedure information, and 4) design part-task practice (Van Merriënboer, 2019). This model describes a scaffolding approach to training, where the less confident a trainee is, the more training support is needed. As the trainee's knowledge and confidence grow, the practice becomes more advanced, and less support is provided. This is broken out by part-task practice where segments of the overall task are practiced in a progressive (i.e. scaffolding) approach.

The SEEV model is another model considered in this paper which is typically used to analyze when an operator is responsible for multiple task execution using a visual interface. This model is built on the premise that attentional allocation is either top-down or bottom-up (Rheem et al., 2021). In 2003, Wickens et al. presented research focused on pilot performance and the application of attentional models to eye-gaze dwell time in simulated piloting tasks (2003). Regarding the SEEV model, attention is guided by Salience, inhibited by Effort, and attracted by Expectancy and Value, as depicted in Figure 4. Using this model, the authors explain how participants react to visual cues related to the task that they are performing. In addition to Wickens' (2003) work, Rheem et. al (2021) has developed a GUI-based tool for SEEV model simulation that could integrate the eye-gaze data into a usable platform to predict pilot attention during tasks. The use of this software could be a possible avenue for future quantitative evaluation of the SEEV model.

We utilized this model in a pilot study assessing gaze and periphery in pilots of differing experiences. The action/response of the user can be correlated to the effort or expectancy that occurs, and the user (in the case of an expert) will intuitively choose how to address a certain task. The previous training and experience of expert pilots were apparent in the detection of objects in the periphery and scanning patterns. Experts demonstrate a high level of expectancy due to their training and a higher level of situational awareness whereas novice subjects will not have this type of intuition.

As technology evolves, the quest to improve our training transfer in terms of time, efficiency, and effectiveness through the use of user experience-based learning tools is proving more viable. In this study, we aim to validate the use of VR immersive technology for specific task training. Further, we will assess the current training methodology and address the need for a designed framework (Gouedy, in preparation).

METHODS

Hypothesis and Variables

Based on a comprehensive literature review, we hypothesize that pilot trainees undergoing training with VR simulation experiences will achieve 'proficiency' standards with a reduced number of in-air task minutes and/or fewer sorties compared to historical student trainees who did not have access to this training technology.

Table 1. Dependent Measures with Definition

Variables	Definition
Sorties to Proficiency	The number of sorties (flights) it takes a pilot to reach the level of 'proficient' competence level.
Minutes to Proficiency	The number of minutes a pilot spends on air refueling attempts (practice) to reach the level of 'proficient' competence level.

Proficiency is defined as the pilot's ability to achieve and maintain end-of-course standards in terms of sustained contact time with fewer than the allotted number of inadvertent disconnects. If the hypothesis is proven accurate, we

will then be able to build out a concrete training model to use incorporating part-task practice for pilots including a regimented VR training approach. This could be extended to other tasks and even basic flight training in general.

Participants

Test subjects were voluntarily recruited from the B-52 Formal Training Unit (FTU) at Barksdale AFB. Training data was collected on 198 pilots trained to fly the B-52 at Barksdale AFB. Of this pool of subjects, 152 of the pilots' participants' data was from historical records of most recent classes without the use of VR supplemental access, while 46 pilots were trained with access to the VR training simulation program. The pilot subjects were assessed based on two different groups: the total control group and the test group. The control group was created out of the historical data of pilots trained with no structured guidance and learning objectives without immersed VR trainers. The data came from secondary data such as de-identified performance grades and program progression. The test group was defined as pilots who were trained with differing levels of VR access and training objectives throughout the formal program. Data for this group included both secondary data comparable to the control group, as well as specific quantitative and qualitative data derived from eye gaze data, questionnaires, and surveys.

Apparatus

In our test group, pilots trained using a VR sled and EARL (Experimental Air Refueling Lightweight) flight-task simulator designed specifically for the Air Force to train air refueling. The sled package includes a Varjo brand headset with embedded eye tracking technology. This immersive experience also integrates the throttles and yoke, in addition to the Varjo base computer software. The software works in tandem with the Varjo headset to capture and record each pilot's flight runs and gaze data. To ensure high granularity in imagery, a computer equipped with a 3080 graphics card was used throughout data collection and assessment.

Procedures

Historical data on current and previously graduated pilots was provided to build out historical metrics on proficiency timeline based on the number of minutes pilots practice air refueling as well as the number of sorties flown. Overall training times and sorties were collected as well as the dates and instructors of each sortie. The class averages of the test pilot groups were used as a comparison to the historical averages of the control pilots.

Each of the groups, whether within the test or control group, completed the course objectives and requirements outlined in the governing

Figure 1. VR Training (left seat) Setup

syllabus for the FTU program. There were three different iterations of syllabi over the span of our data. The differences relevant to our research defined the expectation of minutes where air refueling (AR) was attained during each of the four blocks of the training program. The final class included in the test group incorporated an unofficial training framework that followed a scaffolding concept with skills, practice, and application, using the VR simulation.

Once the aforementioned FTU data was compiled, a statistical analysis was conducted on program performance. The data was first assessed by looking at the entire data set of pilots (152 non-VR-trained pilots vs. 46 VR-trained pilots), across all three iterations of the syllabi. A second assessment (see Table 3) was conducted, exempting pilot classes that graduated under the oldest syllabi requirements of the three. This assessment resulted in an equal sample set of pilots for both the test and control groups (46 non-VR-trained pilots vs. 46 VR-trained pilots). Regression models

were then used to determine the relationship between the dependent (AR time to proficiency and number of sorties to proficiency) and independent variables (VR use and time between sorties).

Analysis was then conducted to assess the cognitive load on the pilots using visual heat maps and the NASA-TLX. The NASA-TLX asks six questions through a digital application that measure pilots' cognitive load at four different points in the training program (see Table 2). The overall scoring calculation of the NASA-TLX surveys identifies the range of cognitive load a pilot experiences during the task execution in each environment.

Table 2. NASA TLX Collection Points

VR Simulated Flight	Sortie in the Jet
Following First attempt	Following First attempt
The subsequent VR sim following in-air proficiency grade by Instructor	Following in-air proficiency grade by Instructor

The eye gaze data was collected from the Varjo headset at the pilot's first VR simulation attempt and a VR simulation attempt following the attainment of a proficient grade in the jet. The Varjo headset is used as part of the B-52 Formal Training Unit (FTU) normal training protocols. Gaze data from the Varjo headset was analyzed using iMotions to observe the focal areas of interest throughout the task activity and to cross-check data from the NASA-TLX regarding the cognitive load of our pilots at different points in the training process and in different environments.

Lastly, subjective pre and post training questionnaires for two of the VR trained classes were administered and reviewed. Subjects completed a pre-training questionnaire at the beginning of the program to capture demographics, pilot background/flight experience, and subjective experience questionnaire. Subjects also completed a post-training questionnaire to capture subjective experience and suggestions for improvements with the training protocols and equipment. Our research team also took note of oral interactions with subjects and instructors throughout the data collection process. These interactions were not formal in nature but rather served as natural and subjective responses during the task.

RESULTS

Historical training data on pilot trainees' flight time (in minutes) and the number of sorties (flights) to proficiency was analyzed to conduct a comparative analysis in SPSS between the test and control groups. The test group utilized differing levels of structured integration with the VR training technology and training model. A t-test was performed to assess whether the differences in each test groups time and sorties are statistical significance. All indicating whether the independent

Table 4. Data Collection Totals

Data Collection Totals	
Total Students	198
Total Minutes of AR Training	64408
Average minutes per student	325.29
Highest # of minutes to training completion	579
Lowest # of minutes to training completion	152

variables, VR as a training tool, has any significant impact on our dependent variables and if so to what extent. A normality test was also conducted on the data sets collected for the # of sorties and minutes until proficiency.

Thirteen groups of non-VR trained pilots and four groups of pilots who had access to VR as a supplemental training tool were assessed for each question related to the efficacy of VR. On the question of whether VR use reduces the number of minutes to reach proficiency, we used the raw data set found in table 4 below. The entire dataset included a total of 198 instructed pilots, all pilots were being trained in the Formal Training Unit at Barksdale Air Force Base on the B-52. We assessed arguably the most challenging task of training air refueling. Over 64,000 minutes of air refueling training time was logged in total. The average applied training time per student was approximately 325 minutes, with a high of 579 minutes and a low of 152 minutes. We did determine that time to proficiency, in minutes,

was not the best metric by which to measure efficacy. The time that a trainee is afforded in applied training in the jet is reliant on secondary factors outside of the control of the instructors. These factors include weather, maintenance issues with either the jet or the tanker, and scheduling conflicts with the tanker that are also impacted by all the aforementioned factors.

The overall results are summarized in Table 3. Each dataset was sanitized before analysis to ensure all subjects remained unidentifiable. Descriptive statistics were calculated for each test group, followed by a *t*-test to determine statistical significance. A quick glance of the results indicates that our pilot trainee groups who received access to the VR tool observed a reduction in both sorties and average time to proficiency, which is statistically significant.

Table 3. Results Summary Table

Research Question	N	Mean	*p* value	Effect Size	Results
Does VR use in training reduce the *number of minutes* to reach proficiency?	Non-VR=152	250.25 min	.032	0.313	Statistically significant, Small Effect Size
	VR=46	227.67 min			
Does VR use in training reduce the *number of sorties* to reach proficiency?	Non-VR=152	9.16 sorties	<.001	0.608	Statistically significant, Medium Effect Size; VR use for training air refueling skills has a measurable positive impact on the trainees.
	VR=46	7.80 sorties			

Proficiency Data

The minutes of training it takes pilots to reach proficiency without VR and with VR were analyzed. A significant difference was measured. Pilot classes that did not have access to VR as a training tool took longer to reach proficiency (M = 250.25, SD = 71.8) than did pilot classes with access to the VR trainer (M = 227.67, SD = 73.54), *t*(196) = 1.858, *p* = .032. This indicates that the students with access to VR were able to reach proficiency sooner than students who didn't. We are 95% confident that the average difference in the number of sorties lies between -1.38 and 46.54. Figure 2 shows the trend at which each pilot class reached proficiency. On average, VR-trained students reached proficiency in 22.58 minutes less than their non-VR-trained counterparts.
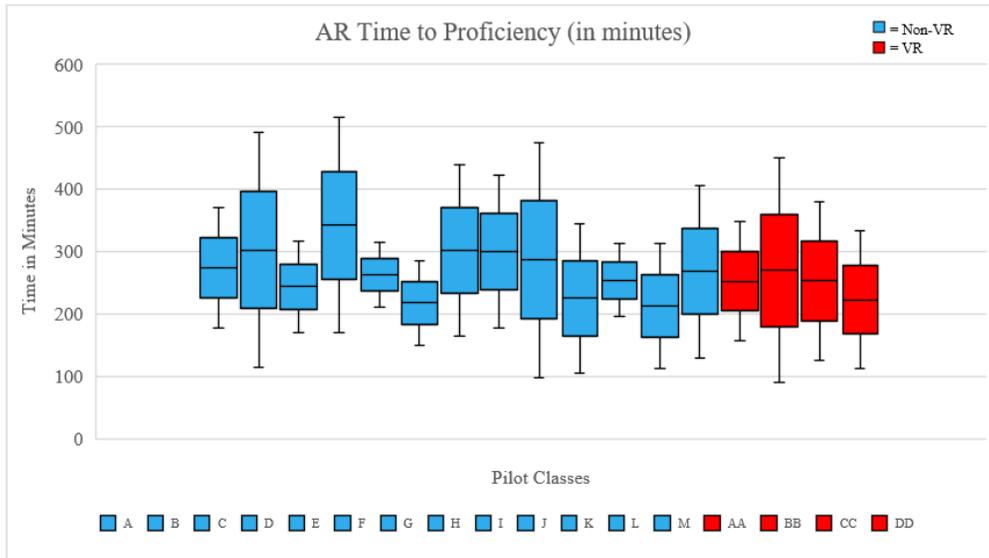
Figure 2. Air Refueling Time to Proficiency (in minutes)

The number of training sorties it takes pilots to reach proficiency without VR and with VR were analyzed. A significant difference was measured. Pilot classes that did not have access to VR as a training tool took longer to reach proficiency (M = 9.16, SD = 2.28) than did pilot classes with access to the VR trainer (M = 7.80, SD = 2.01), $t(196) = 3.617$, $p = <.001$. This indicates that the students with access to VR were able to reach proficiency sooner than students who didn't. We are 95% confident that the average difference in the number of sorties lies between 0.615 and 2.09. Based on our results VR-trained students required 1.35 fewer sorties to reach proficiency compared to their non-VR-trained peers.
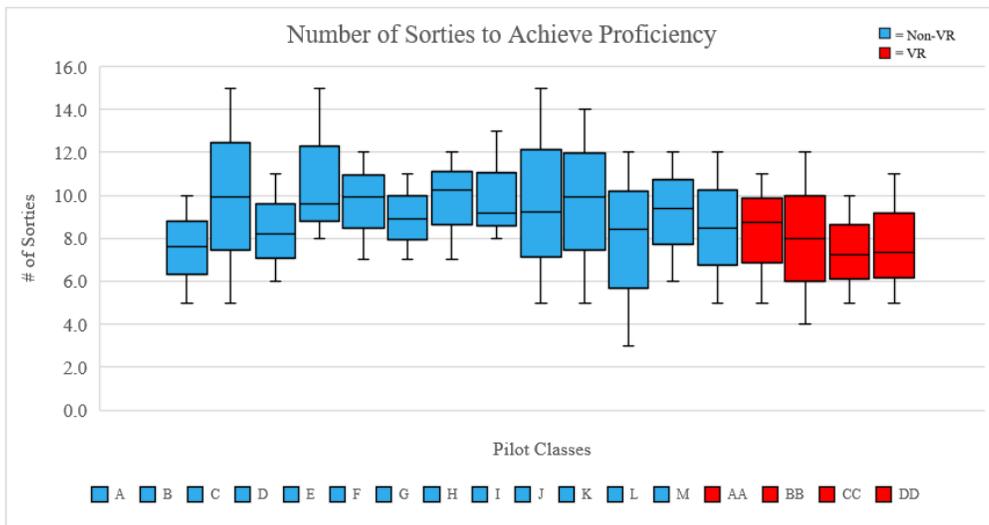

Figure 3. Number of Sorties Until Proficient

Normality was verified by the Shapiro-Wilk test for both the number of sorties to proficient and minutes to proficient and the results indicated to no apparent violation of the assumption (# of sorties to proficient $p = .002$ and minutes to proficient $p < .001$).

A closer look at the VR-trained classes shows a consistent trend in the reduction of the number of sorties it took to reach proficiency. These numbers decrease as the structures and implementation of the VR equipment with the course instruction as students transitioned between the classroom to the flight line increased. This trend supports our effort for the development of a training design that will seamlessly meld the conventional syllabus with VR training tools.

Table 5. Average # of Sorties to Proficiency

Training Groups	Average # of Sorties to Proficiency	Reduction in Sorties from the Historical Average
Historical Average (A-M):	9.28	
Class AA average	8.73	-0.60
Class BB average	8.00	-1.28
Class CC average	7.25	-2.03
Class DD average	7.33	-1.94

Cognitive Load

I)   NASA TLX

As discussed previously, the task of air refueling is a cognitively demanding activity on the pilot team. The NASA Task Load Index (NASA-TLX) survey allowed our team to put a quantitative measure to the movement of cognitive load by score. By looking at the scores of the pilots at several points throughout the training program we are able to observe trends from differing times during the training program. We measured whether the pilot was considered low-medium, medium-high, or high-very-high cognitive load performing air refueling. Table 6 illustrates the mean (average) scores of all students in group DD. We compared three different points: the first sortie to proficient sorties, the first VR session to VR post-proficient sorties, and the first VR session to proficient sortie. In each comparative analysis, we observed a decline in the average score from the first point of measure to the second.

Table 6. NASA-TLX Comparison Scores

	1st Comparison			2nd Comparison			3rd Comparison		
	First Sortie		Proficient Sortie	First VR		VR Post-Proficient Sortie	First VR		Proficient Sortie
Mean	**63.3327**	>	**58.0555**	**61.8053**	>	**34.9997**	**61.8053**	>	**58.6111**
St.D.	17.6836		17.5737	9.9179		13.8816	9.9179		15.1548

II)   Eye Tracking Data

The second data point we used to measure cognitive load was derived from observing eye-tracking data metrics captured during VR flight simulation recordings. Each pilot from class DD completed 5-minute evaluation flight simulations at two points in their training; first VR attempt and post proficiency in the jet.

For this analysis, we focused on three different ocular measurements captured using the Varjo XR-3 headset and analyzed in iMotions computer software; fixation count, fixation duration, and saccade count. The ocular data allowed us to assess the differences between these measurements at the first VR use and post-proficiency attention allocation also observed in NASA-TLX, 2nd Comparison (see Table 6). Mean measurements for fixation count increased from $\mu = 190.6$ at the 1st VR us to $\mu = 231$ at the post-proficient measurements. Likewise, mean measurements for fixation

duration increased from $\mu = 4057.2$ to $\mu = 5140.1$. Mean measurements for saccade count increased from $\mu = 117.1$ to $\mu = 185.5$. A *t* test was also conducted which showed the current dataset was not statistically significant.

The side-by-side comparison of the attention allocation maps (heatmaps) revealed that the pilots' visual focus was consistently centered around the antenna/white line, and the boom operator at both training points. This consistent focus suggests that the pilots were able to maintain their attention on critical elements of the refueling process. We can however observe in the post-check ride VR a tightening of the visual focus which indicates less drifting of visual attention. By tracking these visual attention patterns, we can identify improvements in the pilots' situational awareness and task proficiency, which are crucial for successful air refueling operations.

Figure 4. Heatmap of First VR (on left) and Post Check Ride VR (on right)

Subjective Questionnaires

The data collected from the pre- and post-questionnaires completed by our VR pilot participants provided interesting insight. Of our VR participants, 96% of the pilots had previous flight training either through the military or private/commercial entities. There was an average of 311 flight hours for participating pilots before beginning the FTU program at Barksdale Air Force Base. Of the pilots with previous flight training, 75% had received training in other programs using VR. These previous training programs were associated with training of the T-6, B-1, T-38, or F-16. When asked how well the pilots thought the VR training simulator would help prepare them for air refueling, using the Likert scale with 1 being 'not at all' and 5 being 'fully prepared', results indicated 92% of them responded with 3 and above, as seen in Table 7.

Table 7. Pre-Training Questionnaire Results

How well do you think the VR training simulator will better prepare you for the actual task?				
1	2	3	4	5
0%	8%	44%	28%	20%

In the post-training questionnaire pilots were asked to provide feedback on their experience using the VR trainer in the B-52 FTU Program. When asked if they would recommend the use of the VR training simulator as a training tool for air refueling, 100% of the pilots answered 'yes'. Pilots identified the VR program as being the most helpful in learning the visual references and practicing the correct input actions in conjunction with those reference points. However, the consistent thread when asked about the disconnect between the VR and in-the-jet experience was the power inputs currently built into the VR program.

DISCUSSION

Based on the results we can infer that the opportunity the VR simulation program offers for pilots to practice air refueling outside of the jet helped reduce the number of sorties and overall time needed to become proficient. The results regarding the pilot's reduction in mental workload is promising though inconclusive. The NASA-TLX results indicated a decrease in workload across the points of comparison, which could be attributable to the use of VR as a training tool. Two of the eye-tracking metrics agreed with the NASA-TLX (fixation count and saccades), indicating a decrease in cognitive load (Marquart et al., 2015), which the fixation duration did not. However, this metric is consistent with the findings from Liu et al. (2022), indicating an increase in perceptual load, which makes sense given the nature of the air refueling task. This suggests that incorporating VR as a training tool allows pilots the ability to familiarize themselves with complex tasks in a controlled environment, ultimately enhancing their performance and efficiency during actual flight operations. Additionally, the reduction in mental load could lead to improved decision-making and overall mission success, highlighting the value of integrating advanced training technologies into pilot training programs.

Efficiency of training resources is a key driver behind the implementation of technology-based training environments. However, it is essential to discuss whether VR as a training tool is more successful as a supplement to traditional training design or as a replacement. In our research, we found multiple instances where, even with the VR training tool in place, there was immense value in having a live human instructor on hand with the student during task simulations. This raises the question of whether this need is specific to the rigors of air refueling or if it is something that would be observed in other tasks, both military and civilian.

The use of VR for training provides a level of accessibility and affordability that conventional supplemental training platforms do not. Pilots in the FTU program at Barksdale Air Force Base utilize a WST Simulator created by Boeing/Lockheed Martin, for training purposes. The origin date of these large-scale pieces of equipment is not noted in the 2020 Military Simulator Census, though it has had multiple updates over the years and has been a staple in the training syllabus (FlightGlobal, 2020). These life-size simulators provide a more physical sense of realism that is just now catching up in independent headset units. Though it is important to note that WST Simulator is not designed with the capabilities to practice air refueling. However, updates and improvements to the large simulators are far more costly and time-consuming than a computer programming update typically required with VR programs. VR, with its smaller footprint, allows for greater accessible, shorter and more specific training sessions.

Observing the fluctuation in the mean measurements within selected areas of interest (AOIs), we identified the mental workload shift from the two points in the training program. Considering both perceptual and cognitive load being present to accomplish the air refueling task we should see movement from the 1st VR use measurement to the post-proficient measurement as indicated in Table 7. It is important to note that the small sample size in this current analysis underscores the need for continued collection and a larger sample analysis.

CONCLUSION

Ensuring our airmen are equipped with the best training available is imperative for mission readiness. Real world consequences come into play as we look for ways to ensure our defense programs are able to send airmen into the field as soon as possible and better prepared. However, safety, cost and time of training new pilots continue to be factors. That is where the use of technology-based training tools can play a role. Based on Government Accountability Office numbers, it costs $88,000 for one hour of B-52 operation (U.S. Government Accountability Office, 2022). This cost is similar, whether applied to training mission or defense mission. Based on our assessment, the reduction of the number of training sorties it takes to train airmen can have a significant fiscal impact on the force and the command. If our trends continue with an average 1-2 sortie reduction per student, per class, this could result in an average cost reduction of $10M per training class. While the potential cost impact is impressive, the confidence this resource provides to the airmen observed is invaluable.

REFERENCES

Airforce Technology. (2013). *The world's biggest aerial refueling aircraft*. https://www.airforce-technology.com/features/featurethe-worlds-best-aerial-refuelling-aircraft/?cf-view

American Psychological Association. (2018, April 19). *American Psychological Association*. https://dictionary.apa.org/cognitive-load

Ayers, R. B., & Hamilton, B. A. (2022). A Multi-method Learning Framework for Multi-capable Airmen. *In Proceeding of Interservice/Industry Training, Simulation, and Education Conference*, 1–11.

Blow, E. W. (2023). *Navy training squadron set to fully adopt modernized flight program*. https://www.dvidshub.net/news/437792/navy-training-squadron-set-fully-adopt-modernized-flight-program

Budgeon, M. K., Wolf, B. S., Merkle, M. T., & Senft, D. C. (2022). Quantitative Analysis of Virtual-Reality Training Effectiveness for Cockpit Procedures. *2022 Interservice/Industry 2Training, Simulation, and Education Conference (I/ITSEC)*, 1–12.

Chapman, K. (2023). *US NAVY T-6B UNIT PREPARES TO TAKE LEAD ON "PROJECT AVENGER."* https://www.key.aero/article/us-navy-t-6b-unit-prepares-take-lead-project-avenger

Davis, J. (2024). *B-52 Stratofortress: Raining Down Hate since 1954, and Getting Better with Age*. Intergalactic: An Aerospace Company. https://ig.space/commslink/b-52-stratofortress-raining-down-hate-since-1954

Firsing, S. (2024, March 11). *Flight training takes off with virtual reality*. General Aviation News. https://generalaviationnews.com/2024/03/11/flight-training-takes-off-with-virtual-reality/#:~:text=EASA%20was%20the%20first%20regulator,VR%20flight%20simulation%20training%20device.

FlightGlobal. (2020). *Military Simulator Census 2020*. https://www.flightglobal.com/download?ac=75350

Garcia, A., & Winer, E. (2022). Blending AR and VR to Increase Situational Awareness during Training. *In Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, *22280*, 1–12.

Gentry, Sms. M. A. (1994). *Background Paper on The History of air Refueling*. https://www.airuniversity.af.edu/Portals/10/AFEHRI/documents/AircraftHistory/gentry.pdf

Gouedy, L., Fendley, M., Wolf, B. (2024) B-52 Pilots in Focus (Manuscript in preparation). *In Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, (paper *24240*)

Hawkins, D. (2019, August 30). Pilot Training Next graduates 14 in learning experiment's second iteration. *Air Education and Training Command Public Affairs*. https://www.af.mil/News/Article-Display/Article/1948666/pilot-training-next-graduates-14-in-learning-experiments-second-iteration/#:~:text=Pilot%20Training%20Next%20graduates%2014%20in%20learning%20experiment%27s%20second%20iteration,-Published%20Aug.&text=Air%20Education%20and%20Training%20Command%27s,29.

Hunter, T. (2023, October 25). *Pilots achieve readiness through virtual reality training*. Seymour Johnson Air Force Base. https://www.seymourjohnson.af.mil/News/Article-Display/Article/3568733/pilots-achieve-readiness-through-virtual-reality-training/

Liu, J. C., Li, K. A., Yeh, S. L., & Chien, S. Y. (2022). Assessing Perceptual Load and Cognitive Load by Fixation-Related Information of Eye Movements. *Sensors*, *22*(3). https://doi.org/10.3390/s22031187

Loft Dynamics. (2023, July 17). *FAA Taps Loft Dynamics' VR Flight Simulators to Advance Aviation Goals*. Loft Dynamics. https://www.loftdynamics.com/news/detail-1/faa-taps-loft-dynamics-vr-flight-simulators-to-advance-aviation-goals/

Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of eye-related measures of drivers' mental workload. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, 2854-2861. doi:10.1016/j.promfg.2015.07.783

Mcardle, J., & Hilmer, E. (2022). *Effectively Integrating Technology into Wargames Effectively Integrating Technology into Wargames*. *22123*, 1–12.

Mulders, M. (2022). Vocational Training in Virtual Reality: A Case Study Using the 4C/ID Model. *Multimodal Technologies and Interaction*, *6*(7). https://doi.org/10.3390/mti6070049

National Aviation Academy. (2023). *What Is Aerial Refueling and How Does It Work?* https://www.naa.edu/aerial-refueling/

Rheem, H., Steelman, K. S., & Gutzwiller, R. S. (2021). SEEV Modeler: A GUI-Based Tool for the SEEV Model Simulation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 675–679. https://doi.org/10.1177/1071181321651276

The Link Trainer Flight Simulator. (2020). Naval Air Station Fort Lauderdale Museum. https://www.nasflmuseum.com/link-trainer.html

Tonhäuser, C., & Büker, L. (2016). Determinants of transfer of training: A Comprehensive Literature Review. *International Journal for Research in Vocational Education and Training*, *3*(2), 127–165. https://doi.org/10.13152/IJRVET.3.2.4

U.S. Government Accountability Office. (2022). *Weapon System Sustainment: Aircraft Mission Capable Goals Were Generally Not Met and Sustainment Costs Varied by Aircraft*.

Van Weelden, E., Alimardani, M., Wiltshire, T. J., & Louwerse, M. M. (2021). Advancing the Adoption of Virtual Reality and Neurotechnology to Improve Flight Training. *In Proceedings of the 2021 IEEE International Conference on Human-Machine Systems, ICHMS 2021*, 8–11. https://doi.org/10.1109/ICHMS53169.2021.9582658

VanMerriënboer, J. J. G. (2019). *The Four-Component Instructional Design Model an Overview of its Main Design Principles*. https://www.4cid.org/publications/

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional Models of Multitask Pilot Performance Using Advanced Display Technology. *Human Factors The Journal of the Human Factors and Ergonomics Society*, *45*(3), 360–380. https://doi.org/10.1518/hfes.45.3.360.27250