An Experimental Comparison of the Effectiveness of Various Levels of Simulator Fidelity on Ab Initio Pilot Training

by

Naomi Paul

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2023

Author’s Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Flight simulators have long been used to great levels of success with studies demonstrating that pilots who first train using a simulator reach acceptable levels of competence in less time than pilots who start in the cockpit of an actual aircraft. However, the high costs of simulators are not economically sustainable for the aviation industry and is one of the barriers to entry for those of diverse backgrounds. To address these challenges, there is hope that medium- and low- fidelity simulators including virtual reality head-mounted displays and desktop simulators may be approved for ab initio pilot training as a supplement to real flight training. Ab initio pilot training is pilot training "from the start", training pilots with no experience until achievement of their commercial pilot's license.

Natural tactile interaction is traditionally associated with effective pilot training tasks. Despite the lack of natural tactile interaction in virtual reality, some tasks, such as procedural tasks, simply require exposure to the aircraft environment. This exposure may be achieved through 3-dimensional simulated models of the aircraft in virtual reality.

My research proposes a between-subjects experiment, with 10 participants per group (30 participants total), to quantitatively address this question by analyzing the improvement of pilots completing simple procedural and aircraft handling tasks using either a high-fidelity flight training device, medium-fidelity desktop simulator, or low-fidelity virtual reality simulator. All participants are student pilots at the University of Waterloo with under 20 hours of flight experience, participated in 5 consecutive days of the study, with training effectiveness evaluated through the improvement in performance from Day 1 to Day 5. This research collected objective flight performance assessed by both a flight instructor and through flight data, as well as subjective rating data regarding mental workload, stress level, and experience of simulator sickness. One-way and repeated measures ANOVA analyses were used to analyze the improvement in participants' performance for ab initio procedural and aircraft handling tasks, comparing three between-subject conditions of using a high-fidelity flight training device, medium-fidelity desktop simulator, or low-fidelity VR simulator.

It was found that virtual reality and desktop simulators are as effective as high-fidelity flight training devices for pilot training of procedural tasks, without increasing the risk of experiencing simulator sickness. However, for aircraft handling tasks, participants training using virtual reality or desktop simulators did not improve to the same degree as those training on the high-fidelity flight training device. This provides evidence for the argument towards approving the strategic allocation of virtual

reality for pilot training in combination with existing simulator methods for handling and other tasks, leading to potential cost savings which make training less expensive and more accessible to future pilots.

Acknowledgements

Primarily, I would like to thank my supervisor, Professor Shi Cao, for his guidance and support throughout the entirety of my research process. His feedback challenged me to continue to expand my knowledge and area of comfort by considering different perspectives and opportunities which I otherwise may have missed.

Second, I would like to thank Dr. Suzanne Kearns and Professor Siby Samuel for agreeing to be on my thesis committee. Their expertise and the feedback and comments they provided helped me to further improve my work.

I would also like to thank Brad Moncion, a fellow graduate student under the supervision of Professor Shi Cao, for his support with data collection and for providing a pilot's perspective throughout the various stages of my work.

I am grateful to the Waterloo Institute for Sustainable Aeronautics (WISA) for the use of and guidance with using the ALSIM AL250 flight simulator. This was invaluable to my experiment which I would not have otherwise been able to complete.

I would also like to express my gratitude towards the students at the University of Waterloo who volunteered to participate in my study. I thank you for taking the time out of your busy schedules to participate in my experiment and allow me to collect the data necessary for drawing conclusions in my research.

Finally, I would like to thank my friends and family. The many discussions with you provided me with the advice and support necessary to complete my thesis and achieve this milestone.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AR	Augmented Reality
CAA	Civil Aviation Authority
CPL	Commercial Pilot License
CRM	Crew Resource Management
DSSQ	Dundee Stress State Questionnaire
EASA	European Aviation Safety Agency
FAA	Federal Aviation Administration
FFS	Full Flight Simulator
FOV	Field of View
FTD	Flight Training Device
HMD	Head-Mounted Display
MCH	Modified Cooper-Harper Scale
MSSQ	Motion Sickness Susceptibility Questionnaire
NASA-TLX	NASA Task Load Index
PPL	Private Pilot License
SAS	Simulator Adaptation Syndrome
SSQ	Simulator Sickness Questionnaire
SSSQ	Short Stress State Questionnaire
SSS	Subjective Stress Scale
TC	Transport Canada
TER	Transfer Effectiveness Ratio
TCCA	Transport Canada Civil Aviation

VR	Virtual Reality
WISA	Waterloo Institute for Sustainable Aeronautics
3D	Three-Dimensional

Chapter 1
Introduction

Prior to the global COVID-19 pandemic, which began in 2020, air travel was increasing in popularity as a method of transportation (Alasim & Almalki, 2021). As a result, there have been growing concerns and research regarding aviation safety, including recommendations to improve training and education within the field of aviation; training and education which has not seen any major advancements since the Federal Aviation Administration (FAA) began gradually allowing the use of flight simulators for training in the 1970s (*FAA Issues New Flight Simulator Regulations*, 2016); simulators which were first developed to enhance flight training in the 1930s ("Usability Testing of a VR Flight Training Program," 2021).

Researchers across the globe have begun conducting studies that suggest virtual reality (VR) may be an effective simulation method for training and education, and hypothesize that in the future, with regard to pilot training, VR may be able to replace, or be used in combination with, traditional simulation methods. These traditional methods include flight training devices (FTDs) (high-fidelity simulators used for teaching and practicing flight scenarios prior to an actual aircraft) and desktop simulators (medium-fidelity simulators developed by connecting physical controls to approved simulation software on a desktop computer, typically used for at-home pilot training).

VR was first presented in 1965 by computer scientist and internet pioneer Ivan Sutherland (Mazuryk & Gervautz, 1996) and has since been researched in depth and developed into a rapidly growing technology. It is a platform that uses a combination of hardware, a VR headset and touch controllers, and computer software, including game development engines, to create a simulated environment in which users place themselves into an artificial world (Alasim & Almalki, 2021). Through VR, users are fully immersed in the virtual world where they have the freedom to move around, interact with 3-dimensional (3D) objects, and may be exposed to various visual and auditory stimuli. In summary, VR may be defined as an interactive and immersive experience into a simulated environment (Mazuryk & Gervautz, 1996). For this paper, the term VR will refer solely to immersive VR systems, that is, systems which use head-mounted displays (HMDs) supporting 3D depth vision.

Multiple studies have demonstrated the effectiveness of VR training in various industries such as education, training, healthcare, military, entertainment, firefighting, and others, with VR expected to expand into more domains in the future (Alasim & Almalki, 2021; Boas, n.d.; Buttussi & Chittaro, 2018). One such domain which is gaining interest among researchers is aviation. This is due to the potential of VR training to lead to decreased cost of and increased access to simulated flight training, while maintaining the level of training effectiveness seen in the currently approved simulators.

As the field of VR for pilot training has a lack of studies directly comparing training transfer, training effectiveness, and differences in performance between VR and existing simulation methods, despite the plausible improvements to training which it may bring (Michelle P. Hight et al., 2022), the aim of this research is to begin to develop knowledge on the use of VR for pilot training through a comparison of VR technology to the existing alternatives: high-fidelity flight training devices and medium-fidelity desktop simulators.

The many benefits which are expected to be associated with VR simulator use in pilot training has led some organizations, such as the Embry-Riddle Aeronautical University (*New US University Flight Training Program Using Virtual Reality Cuts Time To Solo By 30%*, 2022) and Alaska Airlines (Kristin Goodwillie, 2022), to implement VR training in their pilot training programs, despite the lack of research on VR use for pilot training. These programs found that they were able to reduce the time it took students to complete their first solo flight by over 30 percent (*New US University Flight Training Program Using Virtual Reality Cuts Time To Solo By 30%*, 2022) and has aided pilots in developing the muscle memory needed to quickly locate switches, saving hours in higher-fidelity simulators and actual aircraft which can better be used for training flight maneuvers (Kristin Goodwillie, 2022).

1.1 Motivation

Due to the lack of natural tactile interaction in VR simulators, the Federal Aviation Administration (FAA) has not yet certified VR simulations. For FAA certification, the current requirement is for simulation software to be accompanied by certified hardware, including both a cockpit and flight controls (*FAA-Certified X-Plane*, n.d.; *X-Plane 11*, n.d.) (It is important to note that while the FAA is part of the United States of America, there is an agreement for co-operation with the Transport Canada Civil Aviation, TCCA, as well as co-operation between the TCCA and the European Aviation Safety Agency, EASA, (*TCCA and EASA Rulemaking Co-Operation*, 2008)). As such, until proven to be effective, simulation software for VR headsets may be beneficial for providing additional practice and repetition of simulations to trainees but will not count towards pilot licensing or currency criteria, and thus are not able to replace any aspect of existing training using physical simulators.

However, if VR simulators were to become certified, there would be many benefits which align to the three pillars of sustainability within the aviation industry: environmental sustainability, economic sustainability, and social sustainability.

1.1.1 Environmental Sustainability

Environmental sustainability is defined by an aviation sector that "dramatically reduces the negative environmental impacts of aviation," (Waterloo Institute for Sustainable Aeronautics, n.d.). Currently, one of the largest negative environmental impacts of aviation comes from the dependence on liquid fuel, however, as simulators become more effective in pilot training and become approved for training, this may lead to reduced emissions from actual aircraft due to the decreased use of actual aircraft in pilot training.

Additionally, noise is a major environmental concern in aviation due to the high sound levels produced by aircraft engines. This concern is eliminated during training through the use of simulators.

1.1.2 Economic Sustainability

Economic sustainability is defined as aviation infrastructure which supports "the long-term economic growth of the sector," (Waterloo Institute for Sustainable Aeronautics, n.d.). To ensure economic sustainability, the aviation industry must be efficient, affordable, and equitable in its transport of goods and people, while supporting the global economy through the millions of jobs provided in the aviation industry.

In the present, aircraft and the associated high-fidelity simulators are one of the largest expenses that contribute towards the prohibitive cost of pilot training. Ab initio flight simulators, that is, flight simulators designed to train pilots "from the start", such as the ALSIM AL250, can cost CDN $500,000 or more. On the other hand, flight simulators designed to train commercial pilots typically cost between $15-20 million, plus operating costs, maintenance, and yearly licensing fees. While student pilots will likely never find themselves flying these full motion simulators until they begin working for an airline, it is important to note just how expensive these simulators are.

In contrast, VR headsets have a cost that is on average below $1000, with negligible costs per hour of use. The only additional cost is the required training software. Currently, there are companies offering flight simulation software which is compatible with existing commercial VR headsets such as

Oculus and HTC Vive. These include X-Plane 11 by *X*-Plane and Prepar3D (pronounced ‘prepared’) by *Lockheed Martin* which both cost $60 (*Prepar3D Product Overview*, n.d.; *X-Plane 11*, n.d.).

Thus, the price of purchasing and running software on a VR headset is a drastic reduction in cost compared to the price of purchasing and running a FTD, which represents one of the most impactful benefits when considering the potential use of VR for training and education in aviation: the potential of significant cost savings.

1.1.3 Social Sustainability

Social sustainability “depends on building equitable, diverse, and inclusive workforces, communities, and global experiences,” (Waterloo Institute for Sustainable Aeronautics, n.d.). This is an important goal for the aviation industry and there is currently a significant lack of diversity among aviation professionals, particularly among pilots. Student pilots are required to travel to specific training facilities to conduct their flight training, which puts further strain on individuals as they must consider the additional finances required for this travel as well as the additional travel time, which takes away from their already busy schedules.

1.1.3.1 Diversity

Within the aviation industry, and particularly among pilots, there is a lack of diversity. In Canada, a mere 6% of all pilots are women (*Women in Aviation*, n.d.), a statistic which is consistent globally. This under-representation of women in aviation remains a significant problem to this day, as the number of pilot licenses held in Canada continues to be skewed towards the male population, as demonstrated in Figure 1.

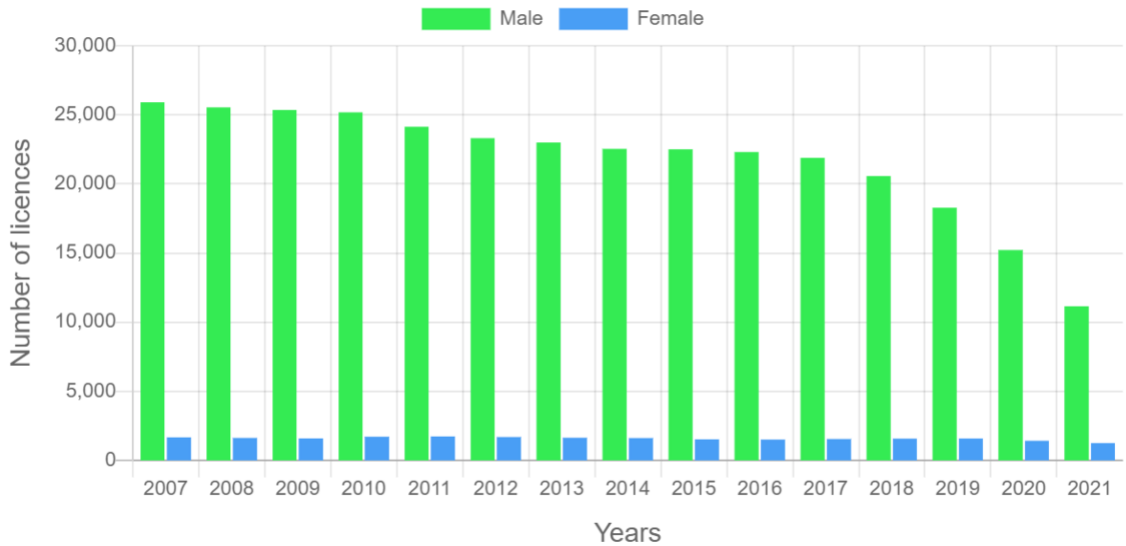


Figure 1: Male-Female Comparison of Private Pilot Licenses Aeroplane in Force (PPL-A) in Canada (*Aviation Personnel Licensing Statistics*, 2022)[1]

In addition to the under-representation of female-identifying pilots, the aviation industry also has a visible lack of diversity with regards to other demographics. Despite the lack of official statistical analysis conducted on the demographics of pilots in Canada, the company Zippia, which democratizes access to data on a wide variety of careers in the United States, has conducted an analysis of pilots regarding demographics such as gender, age, and race. This analysis found that 89.4% of pilots are White, forming a substantial majority, with this remaining steady over the past 10 years, as shown in Figure 2.

[1] Due to the global COVID pandemic, a number of exemptions have been issued since March 2020 to extend pilot's medical validity. This influences the data collected, and the statistics for 2020 and 2021, as such, do not accurately indicate the true number of valid pilot licenses in Canada (*Aviation Personnel Licensing Statistics*, 2022).

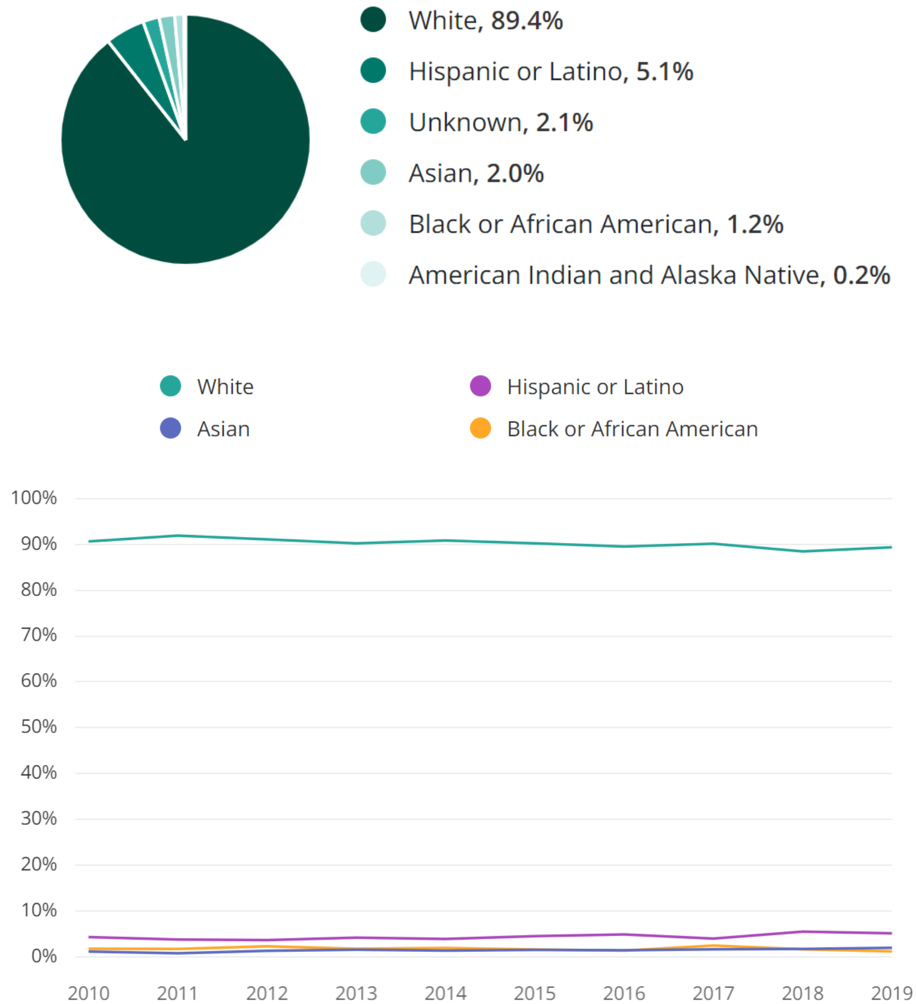


Figure 2: United States Pilot Statistics by Race (*Airplane Pilot Demographics and Statistics in the US*, n.d.)

One of the main factors contributing to this lack of diversity is the high cost of pilot training. In Canada, pilot training, the obtaining of one's Private Pilot License (PPL) and Commercial Pilot License (CPL) costs upwards of $80,000 (*Fees and Financing Your Education*, n.d.). To operate as a professional pilot, one would additionally require their Multi-Engine and Instrument Ratings, which increases the cost of pilot training to $100-125,000.

This prohibitive cost of training means that only those with the financial means necessary can pursue a career as a pilot. Individuals who are a part of a minority, due to their race or other demographics, are more likely to belong to the lower- or working-class and are less likely to be able to afford the prohibitive cost of pilot training than their counterparts. Using VR training to decrease the overall cost of pilot training will thus make training more affordable and therefore more accessible to a diverse group of future pilots.

1.1.3.2 Portability

Commercial flight simulators used in training require approximately 2.7 m x 1.8 m of floor space dedicated solely to the simulator, which may not easily be moved (Wallace et al., 2020). As each plane is unique and may have a different cockpit configuration, a separate simulator may be required for each type of plane used in training, or the controls used in the simulator will be generic and not precisely represent the physical cockpit of the desired aircraft.

In contrast, software for VR headsets allows for the plane configuration to be modified and additional models to be added easily. The only space required for training using VR is a small area clear of obstacles where one may safely enter the virtual environment using a VR headset.

Thus, using VR, student pilots could, in theory, train from anywhere. The VR headset is small enough to fit inside a backpack and simply requires a small space clear of obstacles, the diameter of one’s arm span, to use the headset safely (*Health and Safety Warnings*, n.d.). This means that VR is extremely portable, it is easy to move from place-to-place and may be used almost anywhere, which is beneficial for trainees as it allows them to more easily fit training into their busy schedules as they no longer need to set aside time to travel to the specific locations where the high-fidelity simulators are stored.

Additionally, as VR is much less costly than the full simulators, each trainee may have their own VR headset, which increases accessibility to training as student pilots would have the freedom to train whenever they would like, rather than needing to schedule specific time slots with the high-fidelity simulator.

1.2 Objectives

Research in pilot training is quick to point out the drawbacks of traditional simulators, in particular their prohibitive cost, with researchers increasingly suggesting that the future of aviation will include virtual reality. However, despite the interest in the specific application of VR towards pilot training, there is a lack of studies providing quantitative evidence of the benefits of VR in pilot training. Without quantitative evidence proving the effectiveness of VR in pilot training, demonstrating that VR is just as effective as existing simulators, civil aviation authorities (CAAs) such as Transport Canada (TC) or the Federal Aviation Administration (FAA) will not approve VR use in training due to safety concerns. Safety is paramount in aviation, and as such, any new technology will not be approved for use in training until it has been proven to be at least as effective as current methods.

The overall research objective of this study aims to address gaps in existing literature through an experimental comparison between high-fidelity flight training devices (FTDs), medium-fidelity desktop simulators, and low-fidelity virtual reality (VR) simulators.

Specifically, there are four areas of comparison requiring different measures:

1) Differences in improvement in performance for ab initio procedural pilot training tasks.

2) Differences in improvement in performance for ab initio aircraft handling pilot training tasks.

3) Differences in cognitive fidelity including mental workload and stress.

4) The occurrence of simulator sickness.

1.3 Thesis Overview

The following format outlines the overall structure which this thesis takes:

Chapter 1: Introduction contains the overview, motivation, objectives, and structure of this thesis.

Chapter 2: Background provides an overview of existing literature describing high-fidelity flight training devices, medium-fidelity desktop simulators, and low-fidelity virtual reality simulators including the effect of fidelity on knowledge transfer from simulators to the real-world, relevant aspects of cognitive ergonomics as it relates to pilot training, and the occurrence of simulator sickness. Additionally, existing findings related to the use of simulators in aviation training will be presented.

Chapter 3: Current Study defines the hypothesis and experimental protocol used for this study. The methodology and descriptive statistics of the participants are described.

Chapter 4: Analysis & Results includes a summary of the data collected during the study and important findings.

Chapter 5: Discussion discusses and interprets the meaning of the results.

Chapter 6: Conclusion provides a conclusion, summary of the research, and identifies possible topics for future work.

Chapter 2
Background

The process of becoming a pilot begins with the attendance of a flight school, many of which offer ab initio training to help student pilots train and study for a variety of licenses, including the private pilot license (PPL) and the commercial pilot license (CPL). Each such license and/or rating has unique requirements consisting of medical fitness, ground school instruction, written examination, and in-flight experience (*Qualifications for Pilot Licences and Permits*, n.d.).

Transport Canada does not require student pilots to do any simulator training (or any in class training) before taking their first flight lesson as they are always flying with a certified flight instructor who will teach them about instrument navigation, flight procedures, and general techniques. After their first solo flight, their training becomes a combination of training with an instructor and solo practice. However, student pilots are allowed to count a maximum of 5 of the 45 hours required for their PPL using an approved aeroplane simulator or flight training device, subject to a few guidelines as outlined in Transport Canada's Canadian Aviation Regulations SOR/96-433 (Transport Canada, 2019). This simulation-based training has been proven to be beneficial as previous studies have demonstrated that "pilots who first trained in simulators required less in flight training time before reaching an acceptable level of competence [as compared to pilots who trained without simulators]," (Kaplan et al., 2021).

At the airline level, all training is conducted in simulator devices before ever flying the actual aircraft.

Current approved simulator methods include high-fidelity FTDs, which have been proven to be an effective method of pilot training and at-home desktop simulators which allow for additional practice on-demand, but at a decreased quality compared to the high-fidelity simulators.

In addition to these approved methods, studies are suggesting that VR could be used for pilot training as a low-fidelity simulator alternative as it has the ability to offer a more engaging learning experience than passive training (such as textbooks and posters) (Chittaro et al., 2018), to improve knowledge/skill retention (Chittaro et al., 2018), to reduce the cost of training (Wallace et al., 2020), and to simulate potentially dangerous situations (Wallace et al., 2020). Additionally, student pilots have a tendency to prefer learning methods which allow for the use of theoretical learning to guide their practical flight training (Yi Gao et al., 2013), which aligns well with the appeal of VR in offering an "intuitive manner in which users can control and interact with virtual objects and events," (Hoffman & Vu, 1997) which is thought to reduce cognitive overhead, allowing students to “focus their full attention on the instructional scenario rather than on the semantics of the computer interface,” (Hoffman & Vu, 1997).

2.1 Fidelity

The fidelity of simulators refers to “how accurately a simulator represents the real-world experience,” (Suzanne K. Kearns, 2021) and may be broken down into three elements (Myers III et al., 2018):

1) *Physical Fidelity* describes how closely the simulator replicates the physical aircraft including motion, visual, and sound replication.

2) *Cognitive Fidelity* describes the ability of the simulator to replicate cognitive skills required in flight such as situational awareness, anxiety, stress, and decision making.

3) *Functional Fidelity* describes how closely the actions of the simulator replicate the actions of the physical aircraft.

For this study, the fidelity of the three types of simulators studied are determined mainly by their physical fidelity. There is some influence in terms of the functional fidelity and feedback provided by the simulator to the user. However, as the cognitive fidelity of various simulators is largely unknown, cognitive fidelity does not have a significant impact on the overall definition of fidelity for the simulators studied.

In general, any simulator has benefits which include (Myers III et al., 2018):

1) Providing a safe environment for practicing potentially dangerous scenarios,

2) Reducing the cost of training,

3) Positively impacting the environment by reducing the use of resources,

4) Providing an environment for research, and

5) Allowing for rapid repetitions of different scenarios.

However, along with the many benefits, all simulators have disadvantages. The following list includes many of the disadvantages as well as a short discussion of whether these are manageable or rectifiable (Myers III et al., 2018):

1) Simulator sickness, which can be mitigated through gradually increasing use of the simulator.

2) Inducing adaptation and compensatory skills, which may be mitigated by having the instructor or another individual communicate to the pilots in training the differences between the simulator and the real aircraft.

3) Poor motion cuing, which is a challenge in static or fixed-base simulators, such as those studied, relies on the algorithms used in a particular simulator to align visual input with human motion sensing.

4) Lack of user motivation, which can be mitigated by introducing aspects which motivate users.

5) Complex system architecture, which is a concern when developing new simulators. This is less of a concern when purchasing existing, approved simulators.

6) Over-regulation, which is a challenge in the aviation industry due to a long change process resulting from high levels of concern for safety.

7) Excessive costs of advanced simulators, although these costs are fairly affordable compared to alternatives of training solely on real aircraft.

2.1.1 High-Fidelity Flight Training Devices

Flight training devices are physical devices that artificially re-create the experience of flight, from hardware that includes a model of the cockpit to software that allows you to see the result of interactions with the built-in controls. This allows student pilots to physically interact with aircraft controls, gaining psychomotor skills in a safe and controlled environment.

FTDs, such as the ALSIM AL250 shown in Figure 3, are described as high-fidelity simulators as they closely replicate the physical aircraft being flown, through their design, which includes a full model of the cockpit with all instrumentation and flight controls. They also closely replicate the actions of the physical aircraft, sometimes including platform motions that induce accelerative, decelerative, and turning forces which signal the vestibular system to actually *feel* as if the aircraft is climbing, descending, or turning, enhancing the realism of the simulation through functional fidelity (Drake R. Bradley & Stuart B. Abelson, 1995).

Full flight simulators (FFS) are available in a variety of levels of fidelity A through D. The highest level of FFS are the level D simulators which include a full-scale replica of the aircraft’s flight deck, a six-degrees-of-freedom motion platform, 150-degree field of view, and numerous software models that replicate accurate aerodynamics and environmental conditions (U. S. Department of Transportation and Federal Aviation Administration, 1995). Thus, while the FTD used in this study is referred to as a high-fidelity simulator, it is important to note that there is a wide range of simulator fidelity which exists, including simulators of a much higher fidelity than this FTD.

Figure 3: ALSIM AL250 Flight Simulator at the University of Waterloo

2.1.2 Medium-Fidelity Desktop Simulators

Desktop simulators, such as that shown in Figure 4, are simulators most often used for at-home pilot training (Rongbing Xu, 2022; Rongbing Xu & Shi Cao, 2021). Besides at-home usage, airlines have similar devices, called Virtual Procedures Trainers. They may have low- or medium-fidelity depending on the specific technology and simulator set-up used.

For this study, the desktop simulator has medium physical fidelity due to the attachment of physical flight controls to one's computer setup. This fidelity is lower than that of FTDs as it lacks the full set of physical instrumentation in the replica of the cockpit.

Additionally, the functional fidelity is decreased, mainly due to the lack of motion cues and the confinement of the out-the-window view to what can be displayed on a monitor (Drake R. Bradley & Stuart B. Abelson, 1995). Despite the lack of motion cues, however, visual cues alone are convincing of sensations of motion as nonveridical information depicting movement overpowers veridical information (Drake R. Bradley & Stuart B. Abelson, 1995).

The major differences between desktop simulators and FTDs are that they are less expensive, reducing training costs, and have a lower physical fidelity, lacking some of the natural tactile interaction (defined as physical interaction with flight controls and instrumentation, as well as the feedback resulting from user input) which is necessary for training psychomotor skills requiring physical interaction with the surrounding environment (Jensen & Konradsen, 2018).

Figure 4: Medium-Fidelity Desktop Simulator Setup

2.1.3 Low-Fidelity Virtual Reality Simulators

Virtual reality simulators, such as that shown in Figure 5, run the same software as FTDs and desktop simulators. While VR simulators may have physical flight controls and/or instrumentation attached which increase their fidelity to medium- or high-fidelity, for this study VR simulators are considered to be a low-fidelity simulation method as there is a lack of physical fidelity without the inclusion of physical flight controls and/or instrumentation.

In contrast, the use of a head-mounted display provides a natural 360-degree view, providing higher fidelity than the visual cues of the desktop simulator, with the ability to display the view on a connected computer screen for an instructor (or another necessary individual) to be able to watch the simulated flight.

Figure 5: Virtual Reality Head-Mounted Display with Mirrored Computer Display

The major differences between VR simulators and FTDs are that VR reduces the cost of and space required for training while offering a fully immersive training experience.

Despite these benefits, there has been slow adoption of VR technologies due to the challenges of integrating VR training alongside traditional methods and a resistance to new technology (Lee, 2012). There are also issues of physical discomfort experienced by participants which may include feelings of unsafety due to the lack of access to one’s actual surroundings (Jensen & Konradsen, 2018).

The most important drawback to consider in VR simulations is, however, the loss of natural tactile interaction, which is necessary for one to become familiarized with the flight controls and instruments (Wallace et al., 2020) and to allow for skill transfer of psychomotor skills.

A summary of the comparison of the three types of simulators is included in Table 1.

Table 1: Comparison of Aircraft, Desktop, and Virtual Reality Simulator Fidelity

Type of Simulator	**High-Fidelity Flight Training Device**	**Medium-Fidelity Desktop**	**Low-Fidelity Virtual Reality**
Device and Setup Used for this Study	ALSIM AL250 (*AL250: The Right Device for Your PPL, CPL, % IR/ME Training Needs*, n.d.)	Desktop simulator with 3 monitors, physical flight controls, and instrument panels running X-Plane 11	Oculus Quest 2 with touch controllers, running X-Plane 11
Physical Fidelity			
Physical Aircraft Replication	Full cockpit with physical instrument panels and flight controls, generic to an aircraft type (*Simulators*, n.d.)	Ability to attach physical instrument panels and flight controls, but there is no cockpit (*Desktop Simulators*, n.d.)	A virtual model of the cockpit with simulations of instrument panels and flight controls
Replication of Visual Environment	Seamless 250° x 49° view (*AL250: The Right Device for Your PPL, CPL, % IR/ME Training Needs*, n.d.)	27” (*Desktop Simulators*, n.d.) or the size of the selected monitor; but capable of panning 360°	360° view
Replication of Aircraft Motion	Some have motion platforms with 50° pitch, 60° yaw, 40° roll (*Motion System*, n.d.); else limited to nonveridical information	Limited to nonveridical information	Limited to nonveridical information
Sound Replication	Yes	Yes	Yes
Cognitive Fidelity			
Stress	Generally accepted that psychological and cognitive factors have been replicated, at least to some degree (Dahai Liu et al., 2008)	Does not replicate real-world and physical flight stress (Agata Lawrynczyk, 2018)	Replicates real-world and physical flight stress (Agata Lawrynczyk, 2018)

Type of Simulator	**High-Fidelity Flight Training Device**	**Medium-Fidelity Desktop**	**Low-Fidelity Virtual Reality**
Functional Fidelity			
Aircraft Movement Corresponds to Simulator Input	Some have motion platforms; else limited to nonveridical information (*Motion System*, n.d.)	Limited to nonveridical information	Limited to nonveridical information
Visual Cues Correspond to Simulator Input	Yes (*Visual System*, n.d.)	Yes	Yes
Flight Control Feedback Corresponds to Simulator Input	Physical feedback provided through controls within 20 ms of the input to the control loading system (*Control Loading*, n.d.) Electric control loading of the yoke and rudder to provide realistic flight control forces (AL250: The Right Device for Your PPL, CPL, % IR/ME Training Needs, n.d.)	If physical controls are connected to the simulator, there is some level of physical feedback. If not, there is only virtual feedback.	Limited to virtual feedback
Other Aspects			
Initial Simulator Cost	In general, these simulators may cost $30,000 - $120,000 + ("How Much Does a Frasca Simulator Cost?," n.d.; *Simulators*, n.d.; Wallace et al., 2020) The ALSIM AL250 costs $500,000.	$2700-$8000 + external controls and accessories (*Desktop Simulators*, n.d.)	Headset: < $1000 (*Quest 2*, n.d.; *Vive*, n.d.) Software: $60 (*Prepar3D Product Overview*, n.d.; *X-Plane 11*, n.d.)
Operating Cost	$169 per hour (*The Experience*, n.d.)	Negligible	Negligible

Type of Simulator	High-Fidelity Flight Training Device	Medium-Fidelity Desktop	Low-Fidelity Virtual Reality
Floor Space Required	For full-motion simulators 16’ x 16’ dedicated floor space with 8’ ceiling (Motion System, n.d.); non-motion systems may need as little as 2.7 m x 1.8 m of dedicated floor space (Wallace et al., 2020) The ALSIM AL250 requires 3.7 m x 5.5 m *12.1’ x 18.1;) of dedicated floor space with a 2.5 m or 8.2 m; ceiling (AL250: The Right Device for Your PPL, CPL, % IR/ME Training Needs, n.d.)	Desk or table that is 30” by 27” (*Desktop Simulators*, n.d.)	Small area clear of obstacles, with a radius the size of one’s arm span, for the duration of the training
Level of Immersion	High	Low	High
Current FAA Approval	Yes (*FAA Approval*, n.d.)	Depending on the model, and the software and/or peripherals used (*Desktop Simulators*, n.d.; *FAA-Certified X-Plane*, n.d.)	No (*FAA-Certified X-Plane*, n.d.)
Number of Aircraft Models per Simulator	Generally, a single aircraft type. The ALSIM AL250 models generic single and dual-engine aircraft.	Unlimited software models with generic physical controls (*Desktop Simulators*, n.d.)	Unlimited

Virtual reality, despite being a low-fidelity alternative due to the lack of physical controls and instrumentation, has a higher level of immersion than the alternatives due to the natural 360° view provided by the HMD. Alternatively, the desktop simulator, approved for some training, has the lowest level of immersion due to the limited visuals provided by monitors, however, it has physical controls and instrument panels connected to the simulator. The high-fidelity flight training devices have a full replica of the cockpit; however, these simulators still have a more limited view than VR.

Despite knowledge of these pros and cons for each type of simulator, it is difficult to make an educated decision regarding which type of simulator to use for any given scenario. Typically, the high-fidelity FTDs are used despite their high costs as the lower-fidelity alternatives have unknowns regarding training effectiveness.

Analyzing the performance of these lower-fidelity simulators will provide knowledge towards the impact of these pros and cons on training effectiveness. This better allows the industry to make choices regarding which type of simulator to use based on a cost-benefit analysis and may inform the development of future regulations to approve the usage of lower fidelity simulation for certain training scenarios.

2.2 Skill Transfer

Flight simulators are an essential part of pilot training due to the ability of simulators to provide a safe learning environment for practicing potentially dangerous scenarios and allowing for rapid repetitions of a situation with reduced training costs (Myers III et al., 2018) and reduced environmental impact in comparison to training on an actual aircraft. This training is related to skill transfer; skill transfer being the correct application of a skill learned in a simulation to the real-world counterpart.

It is thought that simulator fidelity, which includes physical, cognitive, and functional fidelity, is closely linked to effective skill transfer, with higher fidelity leading to better skill transfer, and vice versa (Myers III et al., 2018).

When simulators are used for pilot training they are effective at targeting three different types of skills: cognitive skills, psychomotor skills, and affective skills (Jensen & Konradsen, 2018).

2.2.1 Cognitive Skills

Cognitive skills are those skills relating to remembering and/or understanding spatial and/or visual information and knowledge. These types of skills relate to the lower-level cognitive skills as described by Bloom’s taxonomy, shown in Figure 6.

There have not been any research studies which focus on the use of VR in teaching higher-level cognitive skills, but for these lower-level cognitive skills, VR has been found to be beneficial in helping the learner to “remember and understand visual and spatial aspects of a place,” (Jensen & Konradsen, 2018).

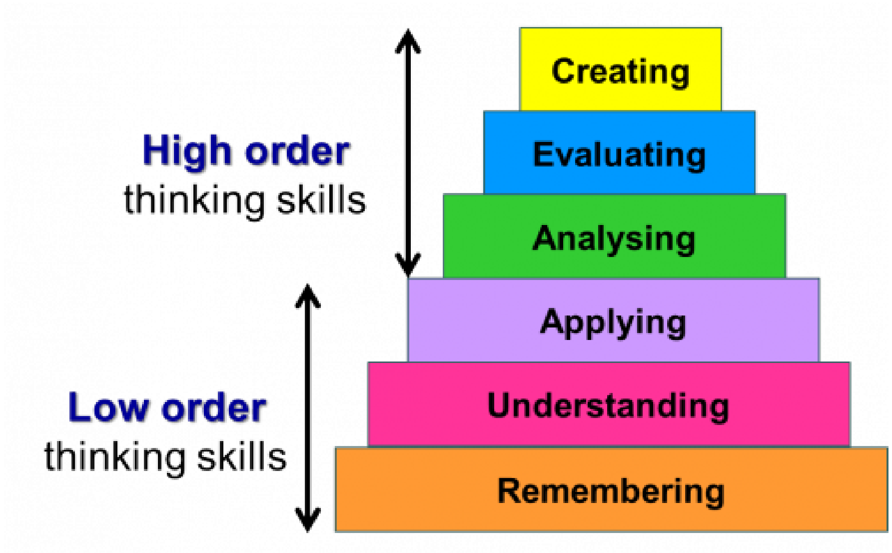


Figure 6: Bloom's Taxonomy of Cognitive Skills

2.2.2 Psychomotor Skills

Psychomotor skills are those skills which are related to movement. There are two primary types of psychomotor skills which must be considered when it relates to pilot training. First, there are psychomotor skills related to the movement of the head, in which VR simulations are highly effective as they allow for the simulation to display a 360-degree view that changes with the user’s natural head movement.

The second type of psychomotor skills required in pilot training are those requiring interaction with the surrounding environment. In this case, skill transfer using VR is not an effective method of skill acquisition without the inclusion of appropriate peripheral haptic and/or tactile devices that allow for natural tactile interaction (Jensen & Konradsen, 2018).

2.2.3 Affective Skills

Affective skills are skills related to controlling emotional responses to stressful and/or difficult situations. Training for affective skills typically requires repetition and highly interactive simulations.

While there is a lack of research covering the potential of VR in the skill transfer of affective skills, it has been suggested that VR may be beneficial in affective skills acquisition provided an emotional response can be created using a combination of sound and imagery in the virtual environment (Jensen & Konradsen, 2018). In other terms, provided that the simulator replicates the cognitive fidelity of the actual aircraft to an acceptable degree, the simulator should be able to aid in the skill transfer of affective skills.

VR provides hope in the transfer of affective skills as the decreased cost of the VR simulator, as opposed to higher-fidelity simulators, allows for the repetition needed within a more reasonable budget, while providing a highly interactive and immersive virtual environment in which the training may occur.

2.3 Flight Training Tasks

One of the major distinguishing features between the three methods of pilot training (high-fidelity flight training devices, medium-fidelity desktop simulators, and low-fidelity virtual reality simulators), besides cost, is the difference in the inclusion of natural tactile interaction.

Natural tactile interaction is the touch sensations and physical feedback given through interactions with flight controls and instrumentation. For example, this may include the feel of the yoke, or the force required to move the yoke when changing the position of the aircraft along the 3-axis (pitch, roll, and yaw). The inclusion of natural tactile interaction when training on a simulator allows students to master the intricate feedback relationship between their control inputs and the resulting changes in both the outside visual environment and the instruments. In understanding this relationship, the student learns to control the aircraft with higher levels of precision (Drake R. Bradley & Stuart B. Abelson, 1995).

Additionally, natural tactile interaction is one of the key elements impacting the fidelity of the simulators. First, the physical fidelity of the simulator is increased with natural tactile interaction because the feel of the instruments and controls are replicated, leading to a higher-quality replication of the real-world physical aircraft. Second, the functional fidelity of the simulator is improved through feedback received from the controllers and instruments, allowing the simulator to replicate the actions of the physical aircraft more closely.

Due to the lack of natural tactile interaction in VR, which is caused by the absence of physical controls and instrumentation connected to the simulations, VR is limited to three key areas in which it best fills the requirements for effective skill transfer. These three areas are:

1) Cognitive skill transfer related to remembering and/or understanding spatial and/or visual information and knowledge,
2) Psychomotor skills related to head-movement, and
3) Affective skills are related to controlling emotional responses to stressful and/or difficult situations as well as non-technical pilot skillsets such as crew coordination, communication, decision making, and situation awareness.

In meeting the requirements for effective skill transfer in these three areas, yet lacking the requirements for effective skill transfer outside of these specific application areas, it is expected that VR performs differently in comparison to existing simulator methods between the two major types of flight training tasks: aircraft handling tasks and procedural tasks.

2.3.1 Aircraft Handling Tasks

Aircraft handling tasks, often referred to as "stick and rudder skills" or "hands and feet skills" are flight training tasks which require pilots to manipulate the primary and secondary control surfaces of the aircraft in coordination to operate the aircraft safely, while respecting speed and structural limitations.

Often, these flight training tasks rely heavily on the movement of the aircraft and thus require motion.

Aircraft handling tasks include flight tasks such as training for an engine failure, steep turn, stall recovery, landing, taxiing, or overshoot.

To include the motion cues necessary for training of aircraft handling tasks, natural tactile interaction in simulators is highly beneficial. As such, when training aircraft handling tasks, higher-fidelity simulators which replicate physical feedback from the aircraft are desired.

2.3.2 Procedural Tasks

Procedural flight training tasks are tasks such as completing procedures and checklists or orientation to build familiarity with the aircraft. Additionally, procedural skills may be summarized by looking at Crew Resources Management (CRM) competencies such as communication, decision-making, situational awareness, and teamwork/professionalism.

These tasks all simply require exposure to the aircraft environment, a cockpit with controls and instrumentation, and do not rely on aircraft movement.

Examples of procedural tasks include flight planning and navigation, pre-flight inspections, and emergency procedures.

As procedural tasks do not require any motion cues or physical feedback from the aircraft, natural tactile interaction in simulators is unnecessary. Thus, VR should be an effective method of pilot training for procedural tasks despite its lack of natural tactile interaction which has typically been deemed necessary for effective skill transfer.

2.4 Evaluation of Simulator Effectiveness

To understand how virtual reality compares to its higher-fidelity counterparts (high-fidelity FTD and medium-fidelity desktop simulators), researchers tend to use a combination of qualitative and quantitative methods to consider performance, cognitive ergonomics, and simulator sickness.

2.4.1 Performance

There are many qualitative frameworks that have been developed by researchers to categorize training outcomes and training evaluation criteria including D. Kirkpatrick's four-level framework, A. C. Hamblin's Model of Evaluation, Peter Warr's CIPO Model, B. R. Virmani & Premila Seth's Model, Peter Bramley's Model of Evaluation, David Reay's Three Phase Model, and Kaufman's Five Levels of Evaluation (Mohanty et al., 2019; Tüzün, 2005).

Additionally, there are models that have been created, such as that by Kunche et al. (Kunche et al., 2011), which focus on the elements that determine training program effectiveness.

However, in aviation, Transport Canada, or the CAA in the country of interest, create flight test guides which detail the skills student pilots must be able to execute and to what degree of competence.

These test guides are used by instructors and evaluators who grade the student pilot's performance on a 4-point scale with a mark of 1 representing observed performance which includes critical errors if the objective is not achieved, 2 representing performance which includes major errors, 3 representing performance with minor errors, and a mark of 4 representing performance that is well executed (Transport Canada, 2021). The grading of these tests is complicated, even to pilots and instructors, but in summary, on a flight test, if you score three or more 1's, you fail your flight test and must repeat the entire exam at a later date. If you score one or two 1's, you partially passed the flight test and are only required to successfully pass the items you previously failed EXCEPT if the total number of 1's and 2's on the grading sheet is greater than five; if you receive two 1's and four 2's, one 1 and five 2's or six 2's, you are still required to redo the test at a later date.

A score of 1 or 2 results in a failure for the specified flight task and the skill would need to be re-evaluated (Transport Canada, 2021).

In addition to these measures, which are qualitative, there are also a variety of quantitative evaluation methods.

One such method, used by McClernon et al., is flight variability (Christopher K. McClernon et al., 2011). In their study, McClernon et al. investigated the impact of stress training on pilot performance during stressful flight operations. They calculated each participant's flight variability across the 15-minute tasks using seven aircraft telemetry measures: pitch, roll, pitch rate, roll rate, lateral acceleration, longitudinal acceleration, and normal acceleration.

Variance indices, such as the flight variability score used by McClernon et al., (Christopher K. McClernon et al., 2011), have been effectively used in prior research as a measure of performance in driving by Mackie & Miller, 1978 and Marcotte et al., 2003, and more recently, McClernon and Miller found that, in aircraft performance, it is "a precise, efficient, and effective index of flight control that describes how smoothly the aircraft is traveling through the air," (Christopher K. McClernon et al., 2011).

2.4.2 Cognitive Ergonomics

Cognitive fidelity describes the ability of the simulator to replicate cognitive skills required in flight such as situational awareness, anxiety, stress, and decision making (Myers III et al., 2018). These are all elements of cognitive ergonomics with established measures to quantify them.

2.4.2.1 Mental Workload

Mental workload is an important consideration as a workload that is too high or too low results in mental fatigue and mental overload, or a loss of vigilance and loss of skill proficiency, respectively. Both lead to subpar training outcomes.

Thus, it is essential that the mental workload is at an appropriate level, which requires researchers to use a tool to quantify mental workload.

In theory, the following equation quantifies mental workload:

$$\frac{Task\ Demand\ (i.e. Mental\ Resources\ Required)}{Capacity\ of\ Human\ Information\ Processing\ (i.e. Mental\ Capacity)}$$

However, in practice, quantifying the number of mental resources required requires complex tools, and there is no standard way of identifying the total capacity of human information processing. As such, tools to quantify mental workload are focused on identifying the mental resources required for a specific task.

There are four key types of tools used to quantify mental workload: secondary task performance measures, subjective ratings, physiological measures, and computational modelling.

The most common tools are subjective measures including the Bedford Workload Scale, the Modified Cooper-Harper Scale (MCH), the Mean Inter-Beat Interval, or the NASA Task Load Index (NASA-TLX) (Mansikka et al., 2019).

Within the aviation domain, the NASA-TLX and MCH are both widely used, and a study by Mansikka et al., specifically related to flight training devices found that both rating scales provide similar results (Mansikka et al., 2019).

The NASA-TLX is more detailed than the MCH, providing an evaluation of mental workload for six scales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Thus, when time is not a concern, the NASA-TLX is preferred, while, for time critical environments, the MCH, which simply rates an individual’s overall experience on a 10-point scale, is beneficial.

The NASA-TLX was developed to improve upon its predecessor, the NASA Bipolar Rating Scale. It includes six dimensions, three related to the demand on the subject (mental, physical, and temporal demands), and the other three related to the individual’s interaction with the task (effort, frustration, and performance).

Each of these six scales are presented to participants on a line divided into 20 equal intervals, anchored by the descriptors Low/High, with the exception being performance which is anchored by the descriptors Good/Poor. These intervals divide the scale from 0 to 100 in increments of 5, with the value rounded up to the nearest 5 if the participant marks the scale between two ticks.

The value on these scales provides an evaluation of the participants mental workload on each of the six dimensions. These six values may be combined into an overall workload scale using weightings determined by each individual participant through a series of 15 pairwise comparisons asking for the most important contributor to workload between the six dimensions.

2.4.2.2 Stress

Flying an aircraft is an inherently stressful situation with pilots experiencing psychological as well as environmental stressors; psychological stressors including aspects such as anxiety, frustration, incentives, etc. and environmental stressors including aspects such as noise, temperature, vibration, light, etc.

Stress has many impacts on individuals including an effect on their affect (feelings, emotions, desires), physiology (heart rate, hormones, etc.), physical performance, and cognitive performance. The Yerkes-Dodson Law is one of the most referenced relationships between stress and performance. This law, shown in Figure 7, demonstrates that stress levels both too high and too low result in performance levels lower than the maximum.

Specifically, when stress is too low, the task is seen as monotonous and boring, resulting in a loss of vigilance, and thus decreasing performance. A stress level that is too high results in feelings of fatigue and exhaustion and can lead to burnout in the long-term, also resulting in decreasing performance. Optimal performance requires that participants experience some level of stress, depending on the task complexity.

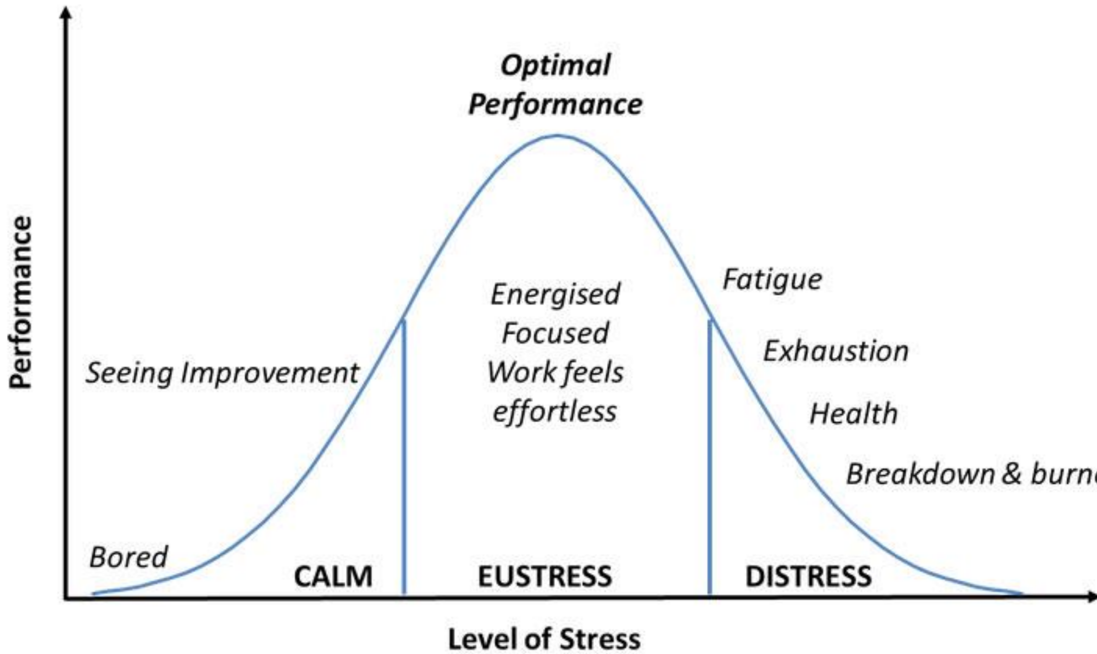


Figure 7: Yerkes-Dodson Law (Hannah Francis, 2014)

Thus, when training in simulators for emergency situations, it is beneficial to replicate these feelings of stress, allowing student pilots to train with higher cognitive fidelity, leading to better performance in real-world situations.

To measure stress, there are two categories of measures: physiological measures such as skin conductance and hormone level, and subjective ratings including the Dundee Stress State Questionnaire (DSSQ) and the Short Stress State Questionnaire (SSSQ).

Although the DSSQ has been proven to be a reliable and valid measure of task engagement, distress, and worry, which are the three broad factors contributing to stress, the major limitation of the DSSQ is its length, being 90 questions, administered both pre- and post-task. To address this limitation, the SSSQ was developed by Helton (William S. Helton, 2004; William S. Helton & Katharina Näswall, 2015). The SSSQ is a shorter 24-item questionnaire to be administered both pre- and post-task, which aligns with the traditional trilogy in psychology: motivation, affect, and cognition.

Despite the SSSQ being an improvement upon the DSSQ, it is still quite lengthy to complete. As such, this study implemented the Subjective Stress Scale (SSS), which is a measure used in a similar study on flight performance conducted by McClernon (Christopher K. McClernon et al., 2011). This measure consists of a single question asked both prior to the experiment as well as after completing the flight tasks, with participants rating their stress level on a 10-point scale; “ten” being the most possible stress that they can withstand before asking to terminate the experiment, and “one” being not stressed at all, like a peaceful, relaxing afternoon.

2.4.3 Simulator Sickness

Simulator sickness is a well-known adverse effect of using VR, oftentimes caused by hardware imperfections which fail to provide accurate stimuli to the user’s senses. However, it is mainly caused by design issues, such as system latency, which cause a discrepancy between the visual motion cues presented by the system and the information sent to and processed by the user’s brain, which causes physical illness as well as frame rate variations where the brain is disoriented when the update does not come at the expected time (Mazuryk & Gervautz, 1996).

Previous research has suggested that the impact of simulator sickness varies drastically with the participants demographic, ranging from symptoms being very rare to almost every participant reporting a symptom. For example, a prior study by Andreoli et al., 2016 and Reiners et al., 2014 found that participants with extensive 3D gaming experience report less symptoms of simulator sickness than those without gaming experience, and a study by Andreoli et al., 2016 found that older participants report more symptoms compared to younger participants (Jensen & Konradsen, 2018).

There are three types of common symptoms experienced by users (Mazuryk & Gervautz, 1996):

1) *Oculomotor dysfunctions* which include effects such as eye strain, difficulty focusing, and blurred vision,
2) *Mental dysfunctions* which include effects such as fullness of head, difficulty concentrating, and dizziness, and
3) *Physiological dysfunctions* which include effects such as general discomfort, headache, sweating, increased salivation, nausea, and stomach awareness.

The symptoms experienced by users may also be short-term (symptoms lasting as little as 5-10 minutes) and long-term symptoms (symptoms lasting longer than 30 minutes, up to 4 hours) (Natalia Dużmańska et al., 2018).

Along with the health concerns associated with the occurrence of simulator sickness, Polcar & Horsejsi, 2015 found that, when simulator sickness presents itself, it skews the learners attitude towards the technology in use negatively, and is correlated with lower learning outcomes (Jensen & Konradsen, 2018).

Due to these impacts of simulator sickness on both the health and safety of learners and the effectiveness of the training, it is essential to have a measure that accurately reflects the simulator sickness experienced by learners.

The most widely used questionnaire to quantify simulator sickness symptoms is the simulator sickness questionnaire (SSQ) published in 1993 by Kennedy et al. (Balk et al., 2013; Kennedy et al., 1993; Walter et al., 2019). Prior studies have proven that the SSQ is robust across all simulator types, making it ideal for studies using multiple forms of simulators.

The SSQ is a questionnaire which asks participants to score 16 symptoms of simulator sickness on a 4-point scale (None, Slight, Moderate, Severe). Kennedy et al., used a factor analysis to determine that these symptoms may be placed into three general categories: Oculomotor, Disorientation, and Nausea, or combined into a single overall score (Kennedy et al., 1993).

The SSQ can be administered both prior to and after the activity of interest with the difference in scores analyzed to remove any impact on the scores caused by symptoms which participants may be experiencing prior to the start of the activity of interest.

Additionally, tools such as the Motion Sickness Susceptibility Questionnaire (MSSQ) may be used to screen participants for susceptibility to motion sickness prior to participation in a study, offering a medium for improving experimental efficacy.

2.5 Existing Literature

There are many methods used for education and training from classroom lectures and reading material such as textbooks to computers and other electronic devices. These methods may be active, getting students involved with the information through hands-on activities involving thinking, discussing, investigation, and creating, or passive, keeping learning an internal process with students listening, note-taking, and asking questions. The method used depends on many factors including one's access to technology, one's willingness to use technology, and the existing technological infrastructure. In our current society where the development and adoption of new technology is rapidly evolving, many have considered implementing virtual reality training. One major appeal of VR for educational purposes is its potential for active learning, coming from the "intuitive manner in which users can control and interact with virtual objects and events," (Hoffman & Vu, 1997) which is thought to reduce cognitive overhead, allowing students to "focus their full attention on the instructional scenario rather than on the semantics of the computer interface," (Hoffman & Vu, 1997).

Previous researchers have suggested that virtual reality methods, when used for education and training, may have many benefits including improving the learner's motivation and enhancing their understanding of content through direct interaction with realism-based practices (Hoffman & Vu, 1997; Lee, 2012; Velev & Zlateva, 2017). Specifically, participants in studies by Bharathi and Tucker (2015) as well as Kleven et al., (2014), as described by Jensen and Konradsen (Jensen & Konradsen, 2018), indicated a moderate preference for VR technology compared to virtual environments created using a desktop simulator. As well, Jensen and Konradsen (Jensen & Konradsen, 2018) reviewed a study by Fernandes et al., (2016), which determined that VR triggered emotions of joy, satisfaction, delight, and enthusiasm. It is important to note, however, that none of the studies measured how these attitudes change over time but focus solely on the participants' initial reactions.

Additionally, one of the most important advantages of VR technology is that it has the potential for aiding users in creating mental models of processes where the tangible counterparts in real life are either inaccessible or non-existent. For example, Hoffman and Vu (Hoffman & Vu, 1997) identify potential use cases where VR can be used to facilitate the understanding of complex 3D processes such as understanding anatomy, molecular structure of force fields, or for phenomena without physical form, such as algebraic expressions.

There have been a number of studies regarding the use of VR for training airplane maintenance such as those by Jamain and Kasirun (Jamain & Kasirun, 2011), Washburn, Stringfellow, and Gramopadhye (Washburn et al., 2007), and Alasim and Almalki (Alasim & Almalki, 2021). As well, there have been studies regarding the use of VR for training airplane passengers on safety procedures including opening of the overwing exits of aircraft in the event of an emergency evacuation (Buttussi & Chittaro, 2021), evacuating the airplane in the event of an emergency ditching (water landing) (Chittaro et al., 2014), evacuating the airplane in the event of a runway overrun (Buttussi & Chittaro, 2018), and life preserver donning (Chittaro et al., 2018).

Outside of passenger safety and aircraft maintenance, there are a few articles worth noting:

First, Brown et al., (2021) performed a literature review which analyzed how VR applications are being used effectively in addressing the ergonomic issues of aircraft crew (Brown et al., 2021). Within this review, they identified that development of VR flight simulators will provide researchers with a method to easily evaluate ergonomic issues related to usability and training. Also, it would provide a method of rapid development for flight deck or cockpit designs without the need to physically build the hardware of all potential designs and display layouts.

Second, Yuviler-Gavish and Gopher (2013) identified six key steps in a task analysis for developing VR training simulators, a modification from traditional task analysis designs (Yuviler-Gavish & Gopher, 2013). These steps include the following:

1) *General definition:* This step is to identify the environment in which the simulator will be used and the main components of the task in the real-world.
2) *Selecting subtasks:* This step requires identifying the subtasks from the task defined in Step 1 that will be simulated in the VR training application.
3) *Define technological requirements for critical subtask elements:* This step focuses on identifying critical components of the subtasks and the technological requirements necessary to meet these components.
4) *Map skills to subtasks:* This step involves identification of the underlying skills (sensorimotor and cognitive) necessary to successfully complete tasks.
5) *Define training requirement:* This step focuses on identifying the requirements needed to achieve the training goals of the simulator.
6) *Integrate technology and training requirements:* This final step is used to evaluate and combine the technological and training requirements into a final list of requirements and constraints that may be used for design recommendations.

Third, G. Alce et al., (2020) studied the potential use of augmented reality (AR) for training pilots on cockpit procedures, called flows (Alce et al., 2020). AR is a virtual environment where the real world is visible along with additional virtual information displayed through an HMD. In this study, a HoloLens was used along with Unity to create an AR version of the paper tiger cockpit simulators; paper tiger's being a cockpit simulation made entirely out of paper, such as that shown in Figure 8. The study found that AR has potential to be a valuable tool for aiding student pilots in learning procedural flows. However, the AR training had limitations due to the narrow field of view (FOV) of AR headsets and the limited number of hand gestures available to the users. To address these downsides of AR training, it was suggested that future studies could be taken to understand if VR headsets may also be a beneficial method of paper tiger training due to the increased FOV, more accurate hand tracking, and higher display resolution, without decreasing the effectiveness of the training through VR's lack of ability to interact with a copilot and the physical operation manual within the simulators.

Figure 8: Paper Tiger Aircraft Model by Flight Vectors (*Cockpit Procedure Trainers - Paper Tigers - Cardboard Bombers*, n.d.)

However, despite the large amount of literature examining VR use in education and training in general, there is a lack of academic studies regarding the use of VR as a replacement for or aid to physical simulators for training pilots, as well as the use of VR for training other aviation personnel including flight crew in cabin positions (e.g., flight attendants and loadmasters) as well as ground crew (e.g., flight dispatchers and marshallers).

One of the few studies in the area focusing on the effectiveness of VR simulations for pilot training is a study conducted by Hight et al., (Michelle P. Hight et al., 2022). This study evaluated the effectiveness of VR simulations, as compared to desktop simulations, in teaching civilian ab initio student pilots. Academic performance data was collected between two groups: one completing the activity using solely the desktop simulations, and the other using VR simulations. The results indicated that students of both groups found the simulations to be enjoyable, but no significant differences were found in the final learning rates between those using desktop versus VR simulators. However, the sample size of the study was quite small, including just 17 student pilots, which combined with the lack of random assignment of participants between groups, limited the statistical analyses between groups.

In conclusion, most studies found that VR may be a beneficial tool for training and education within aviation for a variety of situations including aircraft maintenance, emergency preparedness, and pilot training. These studies suggest that VR should not be used as the sole method of training, but rather used in combination with existing training methods to gain the benefits introduced through VR without sacrificing real-world experience.

However, there is a need for additional researchers to focus on the effectiveness of VR for pilot training to provide evidence towards these claims and demonstrate the benefits of implementing VR use in pilot training, in combination with existing training methods.

Chapter 3
Current Study

3.1 Hypothesis and Study Overview

This study aims to examine differences in the effectiveness of pilot training, as measured through an evaluation of performance, between high-fidelity flight training devices, desktop simulators, and virtual reality simulators for both procedural and aircraft handling tasks.

There are two hypotheses for this study, one regarding procedural tasks and a second regarding aircraft handling tasks. These hypotheses were formed through an analysis of existing literature as well as unstructured interviews with current pilots in the aviation industry, and are defined as follows:

Procedural Tasks: Despite the lack of natural tactile interaction in virtual reality simulators, not all pilot training tasks rely heavily on the use of flight controls and instruments. For such tasks, defined as procedural tasks, which include completion of the before start checklist, it is hypothesized that medium- and low- fidelity simulators, including virtual reality head-mounted displays, will be as effective as traditional high-fidelity simulators for training these skills in ab initio student pilots.

Aircraft Handling Tasks: There are some skills and tasks in pilot training which rely heavily on the use of flight controls and instruments and require trainees to become familiar with the feedback and feel of the aircraft. For such tasks, defined as aircraft handling tasks, which include the execution of take-off and steep turns, it is hypothesized that medium- and low- fidelity simulators, including virtual reality head-mounted displays, will not be as effective as traditional high-fidelity simulators, for training these skills in ab initio student pilots.

To evaluate these hypotheses, participants were recruited for a study which occurred over the course of 5 consecutive days, completing the before start checklist, take-off, and a steep turn in a Cessna 172 simulator: either a virtual reality head-mounted display, desktop simulator, or ALSIM AL250 FTD. Participant’s performance over the course of the week was collected and analyzed to evaluate the improvement in performance due to their assigned training simulator.

This study implemented a between-subjects design, with the type of training simulator to which the participant was assigned, as the independent variable. Dependent variables then included measures of performance (instructor evaluations and flight variability), cognitive fidelity (mental workload and stress), and the experience of simulator sickness.

3.2 Materials

3.2.1 Devices and Tools

Participants in the study were assigned to one of three groups: 1) control group using a high-fidelity FTD, the ALSIM AL250; 2) experimental group 1 using a medium-fidelity desktop simulator; and 3) experimental group 2 using low-fidelity VR simulator.

To ensure consistency between groups, all simulators were set up to model the Cessna 172, a common training aircraft for ab initio pilots. The simulation is based on the Region Waterloo International Airport (CYKF), which contains two runways crossing each other approximately at their midpoints. Runway 08/26 is 7003 feet long; runway 14/32 is 4102 feet long; there are also seven aprons in the airport. For this simulation, pilots took off from Runway 26. Additionally, the weather was configured to be sunny, with dry runways, perfect visibility (i.e., no clouds), and no wind (0 knots heading 0 degrees, meaning static on the ground). With minor fluctuations, the time of day was set to 12:00 PM, with a temperate of 20°C, and air pressure of 29.92 inHg. Default weight, balance, and fuel settings were used in the different simulators to ensure the aircraft was consistent between sessions. Finally, all systems were set to be in perfect working condition with no randomized failures, and the simulation began with engines off.

3.2.1.1 Flight Training Device

The FTD used for this study was the ALSIM AL250 flight simulator operated by the Waterloo Institute for Sustainable Aeronautics (WISA), shown in Figure 9, which is a full simulator with both the physical cockpit and software developed by ALSIM.

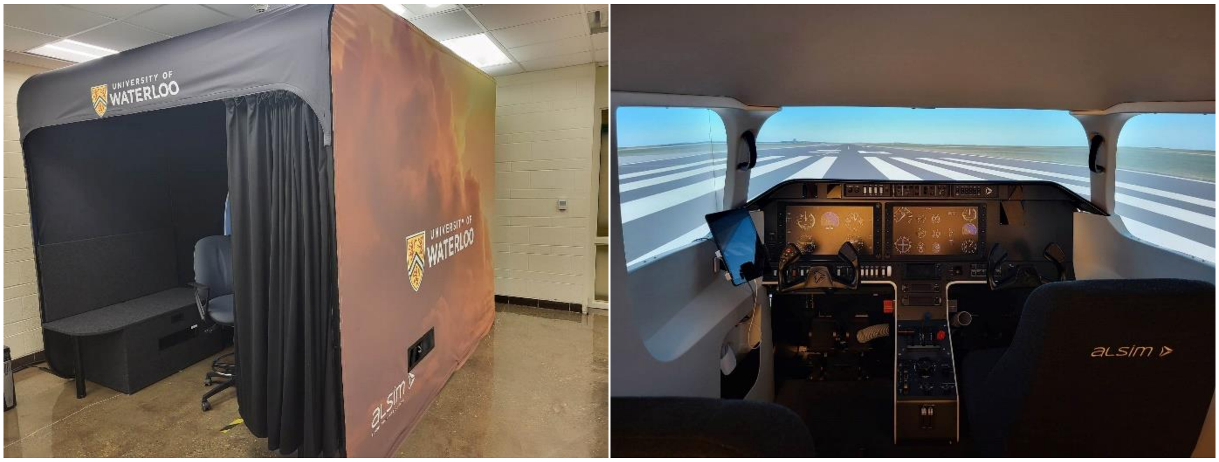


Figure 9: Waterloo Institute for Sustainable Aeronautics ALSIM AL250

The simulator was configured to model a single-engine aircraft, the Cessna 172, with the cockpit configuration shown in Figure 10. It is important to note that the ALSIM device is designed to allow for cockpit configurations including both single- and multi-engine aircraft, and as such, the configuration is a standard layout that does not allow for modifications to precisely replicate the actual Cessna 172 cockpit to a high degree of similarity; the actual Cessna 172 cockpit layout is shown in Figure 11.

Figure 10: ALSIM AL250 Cockpit Configuration

Figure 11: Cessna 172 Cockpit Configuration (Image Source: (*In The Cockpit*, n.d.))

3.2.1.2 Desktop Simulator

The desktop simulator system designed by Rongbing Xu (Rongbing Xu, 2022; Rongbing Xu & Shi Cao, 2021) is comprised of numerous Logitech displays and controllers including a yoke and throttle, rudder pedals, flight multi-panel, radio panel, flight switch panel, and six flight instrument panels, as shown in Figure 12. For this research study, the radio panel and flight multi-panel were disconnected due to computer memory and operating limitations. Additionally, only 4 of the six instrument panels were used to display the airspeed indicator, attitude indicator, altitude indicator, and rpm. Three 24-inch monitors positioned side-by-side display the external surroundings and additional flight instruments found in the cockpit. The base as well as the chair of the simulator are from Volair Sim.

These physical systems were selected to most closely replicate the Cessna 172SP environment, but due to compatibility of the various devices, there remain some differences between the physical cockpit of the desktop simulator and that of an actual Cessna 172, shown above in Figure 11.

Figure 12: Desktop Simulator Cockpit Configuration

In terms of software, the desktop simulator runs X-Plane 11, a flight simulator developed by Laminar Research that runs on both Mac OS and Windows computers, as well as having compatibility for virtual reality. X-Plane is an FAA-certified simulator for pilot training, when combined with certified hardware, and thus is used widely in research and commercial applications.

For this study, the default Cessna 172SP, a single-engine, fixed-wing aircraft, was used, as Cessna is one of the most common training aircraft, as well as to ensure consistency between experimental groups.

3.2.1.3 Virtual Reality Simulator

The virtual reality simulator used for this study is comprised of the Oculus Quest 2 HMD with the included Oculus Touch controllers. The position of the Touch controllers is tracked through sensors embedded in the HMD and contain a series of buttons which allow participants to interact with the virtual cockpit being displayed.

While Meta recently added the ability to use hand tracking to control virtual reality environments in the Oculus Quest 2 headset, X-Plane has not yet incorporated this feature into their application. Until such a time that X-Plane does allow for use hand-tracking technology, it is required that users interact with the application through the Touch controllers.

As the position of the touch controllers is tracked through the HMD, users are able to interact with the virtual cockpit using the touch controllers as if they were their hands. Specifically, users can reach out with the touch controllers until they are hovering over the desired control, at which point in time the control highlights in green, as seen in Figure 13. Users may then press the trigger switch on the touch controller to select, or essentially to ‘hold’ the control. Users then move the touch controller in a way which mimics the real-world controls: for a switch, users will flick their wrist up or down as is moving the switch up or down; for a knob or key, users will rotate their wrist as if physically rotating the control; for the throttle and mixture, users pull or push the controls towards or away from them; and for the yoke, users rotate their hands about the center of the yoke, or push or pull the controls, mirroring the same motions which would be required to control a physical coke.

It is important to note that there are no rudder pedals in this VR condition, and the rudder was not automated by X-Plane. For the flight tasks used in this study, rudder was only required to maintain the centerline during take-off. As an alternative, in VR participants were able to maintain the centerline through control of the yoke.

The Quest headset was connected to a Windows laptop by cable using the built-in Oculus Link software functionality. This connection allowed the simulator software to be run without a dependency on Wi-Fi as well as allowed the researchers to mirror the view of the HMD on the laptop screen, as seen in Figure 13.

In terms of software, the virtual reality simulator runs X-Plane 11, the same software used in the desktop simulator, and as such, is able to model the same default Cessna 172SP as the desktop simulator.

Regarding the cockpit layout, the VR simulation most closely replicated the actual Cessna 172 cockpit configuration, seen previously in Figure 11. The cockpit configuration of the default Cessna 172SP in the VR simulation may be seen in Figure 14.

Figure 13: Virtual Reality Simulator Screen Mirroring onto Researcher's Laptop

Figure 14: Virtual Reality Cockpit Configuration (Julian Lockwood, 2017)

3.2.2 Metrics and Methods of Evaluation

During the study, the cognitive fidelity of the three simulators was assessed through an evaluation of the participants levels of stress and mental workload as higher levels of fidelity correspond with better training outcomes. Additionally, the participants' overall performance using the three simulators was assessed both subjectively and through quantitative data output from the simulators. Finally, the participants' experience of simulator sickness was evaluated as an additional measure due to the potential risk of motion sickness symptoms which occur from simulator use.

3.2.2.1 Stress

Pilots face a variety of psychological and environmental stressors due to the inherently stressful situation of flight. As such, when training, particularly for emergency situations, it is beneficial to replicate these feelings of stress, allowing students to train with higher cognitive fidelity, thus leading to better performance in real-world situations.

To evaluate and measure the stress experienced by participants, as detailed previously in section 2.4.2.2, the Subjective Stress Scale, a single 10-point scale, was used both prior to and after the experiment.

The Subjective Stress Scale, in full details, is included in A.4 Subjective Stress Scale.

3.2.2.2 Mental Workload

Mental workload is another important consideration during pilot training as a workload that is too high or too low results in mental fatigue and mental overload, or a loss of vigilance and loss of skill proficiency, respectively. Both scenarios lead to subpar training outcomes.

The NASA Task Load Index (NASA-TLX) was chosen as the ideal measure for the mental workload experienced by participants. Detailed previously in section 2.4.2.1, the NASA-TLX is an index which provides a measure of mental, physical, and temporal demand, as well as performance and effort (Mansikka et al., 2019).

Full details on the implementation of the NASA-TLX for this study is in A.5 NASA Task Load Index (NASA-TLX).

3.2.2.3 Flight Instructor Assessment

For student pilots, their performance was evaluated through flight tests when working towards various licenses and ratings. On these flight tests, evaluators mark the students' performance using flight test guides developed by Transport Canada.

As outlined in section 2.4.1, these test guides evaluate flight performance for various flight tasks on a 4-point scale (Transport Canada, 2021) with a minimum score of 3 required to pass.

Full details, including a copy of the original and modified test guides may be found in B.1 Flight Test Guide.

3.2.2.4 Flight Variability

Within literature in the field of aviation, flight variability has been found to be "a precise, efficient, and effective index of flight control that describes how smoothly the aircraft is traveling through the air," (Christopher K. McClernon et al., 2011). As such, flight variability was chosen as a quantitative measure of performance during this study.

Specifically, the variance from each participant's flight was used to evaluate four aircraft telemetry measures: altitude, airspeed, bank angle, and heading.

3.2.2.5 Simulator Sickness

One of the major drawbacks of using any form of simulation, and in particular VR simulation, is that participants may experience simulator sickness, which is nearly identical to symptoms of motion sickness. As such, to analyze the occurrence of simulator sickness, the simulator sickness questionnaire (SSQ) (Kennedy et al., 1993), detailed previously in section 2.4.3 was used both prior to and after the experiment. This questionnaire has been proven to be robust across simulator types and is the most widely used questionnaire to quantify simulator sickness symptoms (Balk et al., 2013).

The SSQ used for this study is in A.3 Simulator Sickness Questionnaire (SSQ).

3.3 Participants

To be eligible for the research study, participants were required to be student pilots at the University of Waterloo who have not yet completed their first solo flight and have a maximum of 20 hours of flight experience with any model of the Cessna 172. Student pilots with no flight experience (0 hours) were also welcome to participate. Participants were screened to ensure that they did not have any existing medical conditions that may lead to increased risk of discomfort and other health impacts due to use of a simulator such as susceptibility to motion sickness, experience of seizures, and possible interference with medical devices. The screening questionnaire used for this study, along with associated cutoffs and acceptable ranges, is provided in A.1 Screening Questionnaire.

After signing a consent form, additional data was collected regarding age, gender, flight experience, and familiarity with the Region of Waterloo International Airport (CYKF). The demographic questionnaire used for this study is provided in A.2 Demographic Questionnaire.

Once screened, participants were randomly assigned to one of the three groups (flight training device, desktop, or virtual reality), with some manual reassignments to balance gender, flight experience, and susceptibility to motion sickness.

Participants were solicited through an email sent to the aviation student mailing list at the University of Waterloo.

To statistically compute the necessary sample size, GPower 3.1.9.4 (Franz Faul et al., 2019) was used. The parameters used for the ‘ANOVA: Repeated measures, between factors’ statistical test were $\alpha = 0.05$, $\beta = 0.80$, a large effect size ($f = 0.50$), and two measures (a pre- and post-test). Using these parameters, an a priori power analysis computed the required sample size to be 30 participants.

For the current study, a sample of 30 participants were recruited by email from the University of Waterloo's Aviation program; all participants fully completed the study. The 30 participants comprised of 10 females and 20 males with an average age of 19 (Mean =18.83, SD = 1.90).

When broken down into the three groups, the control group, assigned to the high-fidelity FTD, included 7 Males, 3 Females, with an average age of 19 (Mean = 18.80, SD = 1.23), an average of 3.5 hours of flight experience (Mean = 3.5, SD = 6.96), self-assessed susceptibility to motion sickness (1 being not at all susceptible, 10 being extremely susceptible) of 1.90 (Mean=1.90, SD=.99), and self-assessed familiarity with the Region of Waterloo International Airport (1 being extremely unfamiliar, 10 being extremely familiar) of 2.70 (Mean=2.70, SD=2.06).

Experimental Group 1, assigned to the medium-fidelity desktop simulator, had a higher percentage of females, with 5 Females and 5 Males, although the average age was similar at 19 (Mean=19.20, SD=2.39). The experimental group had an average of 1.6 hours of flight experience (Mean=1.60, SD=3.34), self-assessed susceptibility to motion sickness of 2.30 (Mean=2.30, SD=1.06), and self-assessed familiarity with the Region of Waterloo International Airport of 3.80 (Mean=3.80, SD=1.23).

Experimental Group 2, assigned to the low-fidelity VR simulator, had a similar gender ratio to the control group with 2 Females and 8 Males. This group had an average of 2.5 hours of flight experience (Mean=2.49, SD=3.00), self-assessed susceptibility to motion sickness of 2.0 (Mean=2.00, SD=1.05), and self-assessed familiarity with the Region of Waterloo International Airport of 3.6 (Mean=3.60, SD=2.72).

Using IBM's SPSS (*IBM SPSS*, n.d.) to calculate a one-way between-subjects ANOVA, given that the assumption of homogeneity is true as determined through Levene's test for equality of variance, it may be seen that, despite small discrepancies between groups, there are no statistically significant differences in gender distribution, mean age, hours of flight experience, self-assessed susceptibility to motion sickness, or self-assessed familiarity with the Region of Waterloo International Airport. The calculations for Levene's test for equality of variance, the ANOVA, and the effect sizes may be found in Table 2, Table 3, and Table 4, respectively.

Table 2: Levene's Test for Homogeneity of Variance for Participant Demographics

		Levene Statistic	df1	df2	Sig.
Gender_Num	Based on Mean	2.408	2	27	.109
Age	Based on Mean	1.120	2	27	.341
Hours	Based on Mean	1.671	2	27	.207
MotionSickness	Based on Mean	.332	2	27	.721
WaterlooFamiliarity	Based on Mean	2.230	2	27	.127

Table 3: ANOVA Table for Participant Demographics

		Sum of Squares	df	Mean Square	F	Sig.
Gender_Num	Between Groups	.467	2	.233	1.016	.375
	Total	6.667	29			
Age	Between Groups	2.467	2	1.233	.327	.724
	Total	104.167	29			
Hours	Between Groups	18.074	2	9.037	.395	.678
	Total	636.183	29			
MotionSickness	Between Groups	.867	2	.433	.403	.672
	Total	29.867	29			
WaterlooFamiliarity	Between Groups	6.867	2	3.433	.785	.466
	Total	124.967	29			

Table 4: ANOVA Effect Sizes[a,b] for Participant Demographics

			95% Confidence Interval	
		Point Estimate	Lower	Upper
Gender_Num	Eta-squared	.070	.000	.251
Age	Eta-squared	.024	.000	.159
Hours	Eta-squared	.028	.000	.172
MotionSickness	Eta-squared	.029	.000	.174
WaterlooFamiliarity	Eta-squared	.055	.000	.227

a. Eta-squared is estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Some of the discrepancies between groups, such as the higher proportion of female participants in Experimental Group 1, were due to last-minute reassignments of participants to alternate groups due to illness during the COVID-19 pandemic which shut down access to the ALSIM AL250 temporarily. Despite these last-minute reassignments, the research team did their best to balance the descriptive statistics between groups including age, flight experience, susceptibility to motion sickness, and familiarity with the Region of Waterloo International Airport.

Additionally, some participants completed a few hours of flight training between the time they signed up for the study to the time they began participating. As participants were originally assigned to groups based on the hours listed from the screening, this led to a few discrepancies in the number of hours of flight experience between groups. It is important to note, however, that participants were prohibited from completing any flight training during the week of the study to ensure that their improvement in performance was solely due to their training during the study and not due to external training. The total hours of flight experience for each participant were calculated as the maximum hours listed between the two values.

A summary of the sample's descriptive data calculated using IBM's SPSS (*IBM SPSS*, n.d.) is provided in Table 5.

Table 5: Descriptive Statistics - Means

Group		Age	Hours of Flight Experience	Susceptibility to Motion Sickness	Familiarity with Waterloo Airport	Gender (Male=0, Female=1)
FTD	Mean	18.80	3.500	1.90	2.70	.30
	N	10	10	10	10	10
	Std. Deviation	1.229	6.9642	.994	2.058	.483
	Minimum	18	.0	1	1	0
	Maximum	22	22.0	3	7	1
	Range	4	22.0	2	6	1
Desktop	Mean	19.20	1.600	2.30	3.80	.50
	N	10	10	10	10	10
	Std. Deviation	2.394	3.3400	1.059	1.229	.527

	Minimum	18	.0	1	2	0
	Maximum	25	10.0	5	5	1
	Range	7	10.0	4	3	1
VR	Mean	18.50	2.490	2.00	3.60	.20
	N	10	10	10	10	10
	Std. Deviation	2.014	3.0039	1.054	2.716	.422
	Minimum	17	.0	1	1	0
	Maximum	24	8.0	4	9	1
	Range	7	8.0	3	8	1
Total	Mean	18.83	2.530	2.07	3.37	.33
	N	30	30	30	30	30
	Std. Deviation	1.895	4.6837	1.015	2.076	.479
	Minimum	17	.0	1	1	0
	Maximum	25	22.0	5	9	1
	Range	8	22.0	4	8	1

3.4 Procedure

This study used a between-subject design, assigning 10 participants to each of three experimental groups.

Each participant provided consent before they began participation in the study. Upon providing consent, participants completed a short demographics questionnaire to provide additional personal information not collected through the screening questionnaire (see A.1 Screening Questionnaire). Following the demographic questionnaire, participants began the first of five sessions run consecutively over the course of a week, as shown in Figure 15.

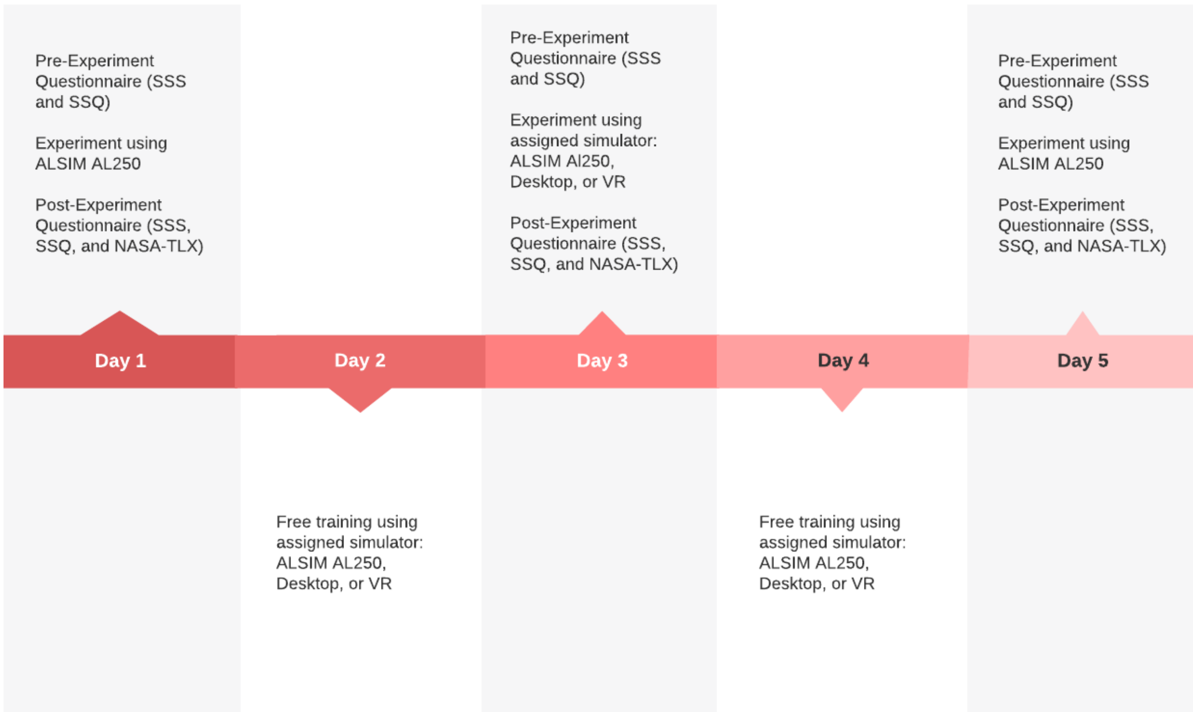


Figure 15: Experimental Procedure

3.4.1 Session 1

During the first session, all participants used the ALSIM AL250 (high-fidelity FTD) as this session served the purpose of assessing participants' baseline knowledge and flight skills. First, participants filled out the SSS and SSQ to assess their baseline level of stress and simulator sickness, respectively. Following the questionnaire, participants received a briefing which provided an overview of the key flight instruments (how to read the airspeed, attitude, and altitude indicators), as well as an introduction to the flight tasks they will completing (before start checklist, take-off, and a steep turn). After the briefing, participants were asked to take a seat in the simulator and were given a few minutes to familiarize themselves with the controls.

In the simulator, participants were asked to follow-along with the before start checklist, which was read to participants verbally, and verbally respond when each item had been accomplished. Checklists were read verbally to match the common two-crew cockpit operations in which one pilot reads aloud the checklist while the other pilot completes the tasks. This also allows for higher levels of consistency between experimental groups as the virtual group is unable to read the checklist while immersed in the simulation. However, due to differences in the cockpit configuration and controls, there are minor differences between the checklist used for the FTD, desktop, and VR groups. These checklists may be found in B.2 Flight Task Checklists.

After completing the checklist, participants were asked to perform a take-off procedure which consisted of releasing the brakes, advancing both the mixture to full rich and the throttle to full open, checking the engine instruments, rotating at 55 knots, and climbing between 70-80 knots. Participants were asked to maintain these parameters for the first 1000 ft of the climb, after which they were given the freedom to make small turns to further familiarize themselves with the simulator and the feel of the controls.

Once reaching an altitude of 4500 ft, participants were asked to perform a steep turn which consists of making a full 360° turn from a heading of North while holding an altitude of 4500 ft, airspeed of 100 knots, and a bank angle of 45°.

After completing these flight tasks, participants were asked to fill out the SSS, SSQ, and NASA-TLX.

3.4.2 Session 2

For Session 2, participants were divided into their experimental groups (FTD, desktop, of VR) and were given a full 30-minutes to practice the three flight tasks (before start checklist, take-off, and steep turns); there were no questionnaires administered.

Typically, this 30-minute session allowed participants to complete the entire procedure twice.

3.4.3 Session 3

Session 3 followed the same procedure as Session 1 with the exception being that participants completed the flight tasks using the simulator assigned to their experimental group: FTD, desktop, or VR.

3.4.4 Session 4

Session 4 followed the same procedure as Session 2.

3.4.5 Session 5

Session 5 followed the same procedure as Session 1, serving as an evaluation of participants' final knowledge and flight skills after the week of training.

Chapter 4
Analysis & Results

For all analyses, a one-way between-factors ANOVA or a repeated measures ANOVA was conducted using SPSS (*IBM SPSS*, n.d.) software for three groups of ab initio student pilots for each type of measure collected.

4.1 Performance

4.1.1 Instructor Evaluations

Instructor evaluations were recorded by an airline pilot for each of 18 criteria, divided into the three flight tasks, based off the Transport Canada private pilot license flight test guide, as follows:

Before Start Checklist

A) Demonstrate an awareness of other persons and property before and during engine start.

B) Use the appropriate checklist provided by the manufacturer or aeroplane owner.

C) Accurately complete the engine and aeroplane systems checks.

D) Check flight controls for freedom of operation and correct movements.

Takeoff

E) Complete appropriate checklist.

F) Check for traffic.

G) Advance throttle smoothly to takeoff power.

H) Maintain directional control during the takeoff roll.

I) Rotate at recommended airspeed of 55 KIAS (+10/-5 knots).

J) Accelerate to and maintain recommended climb speed of 70-80 KIAS (+10/-5 knots).

Steep Turn

K) Perform and maintain an effective lookout before and during the turn.

L) Roll into and out of turns using smooth and coordinated pitch, bank, yaw, and power control.

M) Roll into a coordinated turn with an angle of bank of 45°.

N) Maintain coordinated flight.

O) Maintain the selected altitude of 4500 ft (+/- 100 ft).

P) Maintain airspeed of 100 KIAS (+/- 10 knots).

Q) Maintain 45 degrees angle of bank (+/- 10°).

R) Visually recover from the turn at the pre-selected recovery reference point (+/- 10°).

Participants scores for these 18 criteria were averaged to determine an overall score for each type of flight task: the before start checklist, takeoff, and the steep turn. The improvement in their performance was then defined as *Final performance (Day 5) – Initial performance (Day 1),* giving an overall improvement score for each of the three flight tasks. A repeated measures ANOVA was used to analyze this improvement as a result of the two independent variables: first, a between subject's variable, the type of simulator used for training (VR simulator, desktop simulator, or FTD), and second, a within subject's variable, the type of flight task (before start checklist, takeoff, or steep turn).

First, the tests of within subject's effects found a significant interaction between the type of flight task and the type of simulator used for training ($F(4,54) = 2.915$, $p = .030$, $\eta_p^2 = .178$).

Considering the type of flight task alone, the tests of within subject's effects were not significant ($F(2,54) = .209$, $p = .812$, $\eta_p^2 = .008$), meaning that the degree of improvement from training is not different across the three types of flight tasks.

In contrast, for the test of between subject's effects, it was found that there is a significant impact of the type of simulator used on the improvement in participant's performance ($F(1,27) = 263.449$, $p < .001$, $\eta_p^2 = .907$).

More specifically, using a one-way ANOVA, it was found that there was a significant impact of the type of simulator used for the before start checklist ($F(2,27) = 9.247$, $p < .001$, $\eta_p^2 = .581$) and the steep turn ($F(2,27) = 7.587$, $p = .002$, $\eta_p^2 = .544$), whereas there was not a significant difference for the takeoff ($F(2,27) = .773$, $p = .472$, $\eta_p^2 = .225$).

First, for the before start checklist, this result is surprising as it was hypothesized that participants training on any of the three simulator types would exhibit similar improvements in performance for procedural tasks, such as the before start checklist, whereas it was found that the FTD and desktop simulators led to better performance with the mean improvement being 1.40 and 1.05, respectively, while the virtual reality simulator led to a mean improvement of .575. Running a pairwise comparison, the Tukey test, it was found that, specifically, the difference between these mean improvements is non-significant between the FTD and desktop groups (p = .182), on the cusp of being significantly different between the desktop and VR groups (p = 0.51), and significantly different between the VR and FTD groups (p < .001). To further analyze this difference between the hypothesis and results, a further one-way ANOVA was performed to understand for which specific criteria the three groups improved similarly versus different.

For the before start checklist, the FTD group demonstrated significantly greater improvement with a large effect size for criteria A (F(2,27) = 7.038, p = .003, η_p^2 = .343) and B (F(2,27) = 4.500, p = .021, η_p^2 = .250), while all groups improved similarly for criteria C (F(2,27) = 2.163, p = .135, η_p^2 = .138) and D (F(2,27) = 3.000, p = .067, η_p^2 = .182). This discrepancy between the hypothesis and results is expected to have been caused by the high-fidelity simulator's lack of resemblance to the actual Cessna cockpit, which was accurately represented by the VR simulation. As such, the VR group became familiar with the layout of an actual Cessna 172 cockpit, while their performance was evaluated upon returning to the less-accurate layout of the ALSIM AL250 simulator.

Second, for the takeoff, this result is also surprising as it was hypothesized that participants training on the virtual reality simulator and desktop simulator would not improve to the same degree as participants using the higher-fidelity FTD for aircraft handling tasks, such as takeoff, whereas it was found that the FTD, desktop, and virtual reality simulators led to similar improvements in performance with the mean improvement being 1.00, 1.00, and .817, respectively. A pairwise comparison provided further insight, demonstrating that the difference in mean improvement between the FTD and desktop groups (p = 1.00), the desktop and VR groups (p = .536), and VR and FTD groups (p = .536), are all non-significant. To further analyze this result, a one-way ANOVA was performed to investigate the specific criteria for which the three groups improved similarly versus different.

It was found that the improvement in performance of the three groups was similar for criteria E (F(2,27) = .692, p = .509, η_p^2 = .216), F (F(2,27) = 2.700, p = .085, η_p^2 = .371), G (F(2,27) = 2.562, p = .096, η_p^2 = .363), H (F(2,27) = .518, p = .602, η_p^2 = .192), and I (F(2.27) = .675, p = .518, η_p^2 = .214). The only criteria for which the virtual reality group demonstrated significantly worse improvement in performance for the takeoff was for criteria J (F(2,27) = 4.380, p = .023, η_p^2 = .214). Criteria E, F, G, and I are more procedural, which explains why the lower fidelity VR and desktop groups were able to demonstrate performance similar to that of the FTD group. As such, the only surprising result is that the VR group was able to demonstrate comparable improvement to the aircraft group for criteria H. This discrepancy between the hypothesis and the results is expected to have been caused by the fact that controller feedback is less important for criteria H, maintaining the centerline during take-off, than it is for the majority of aircraft handling tasks, such as criteria J, and the criteria for the steep turn.

Third, as expected, the improvement for the steep turn procedure did vary significantly between simulator groups, with the improvement for the FT, desktop, and virtual reality simulators being 1.525, .863, and .475, respectively. Again, a pairwise comparison was used to provide further insight, finding that the difference between the FTD and desktop groups (p = 0.56) is on the cusp of being significant, the difference between the desktop and VR groups (p = .344) is non-significant, and the difference between the VR and FTD groups (p = .002) is significant. Despite this alignment with the hypothesis, a one-way ANOVA was conducted to investigate the specific differences in improvement between groups for all criteria included in the steep turn.

This ANOVA analysis found that there were significant differences in improvement between groups, with a large effect size, for criteria L ($F(2,27) = 6.056$, $p = .007$, $\eta_p^2 = .503$), M ($F(2,27) = 3.553$, $p = .043$, $\eta_p^2 = .412$), N ($F(2,27) = 4.846$, $p = .016$, $\eta_p^2 = .464$), O ($F(2,27) = 4.268$, $p = .025$, $\eta_p^2 = .442$), P ($F(2,27) = 3.964$, $p = .031$, $\eta_p^2 = .430$), Q ($F(2,27) = 3.433$, $p = .047$, $\eta_p^2 = .407$), and R ($F(2,27) = 5.253$, $p = .012$, $\eta_p^2 = .478$). The only exception where the improvement in performance of the virtual reality group was close to that of the other groups was for criteria K ($F(2,27) = .048$, $p = .954$, $\eta_p^2 = .041$), which is a visual lookout. As VR has a natural 360° view, it is not surprising that the VR group was able to show comparable improvement to that of the FTD group for this specific criterion.

The improvement in performance for the FTD, desktop, and VR groups for each of the 16 criteria may be visualized in Figure 16. For all criteria, A to R, a one-sample T test revealed that the improvement in performance was significantly above 0 with a significance of $p < .048$ for all three types of simulators with a few exceptions: criteria G for the FTD group has a significance level p = 0.52; criteria O and Q for the desktop group had a significance level of p = .122 and .172, respectively; and criteria D, O, P, and R for the VR group had a significance level of p = .084, .0139, .0278, and .217, respectively.

Criteria O, P, Q, and R, for which the desktop and VR group demonstrated improvement not significantly greater than 0, were part of the steep turn, for which it was predicted that these simulators lacked the natural tactile interaction necessary to successfully train ab initio pilots.

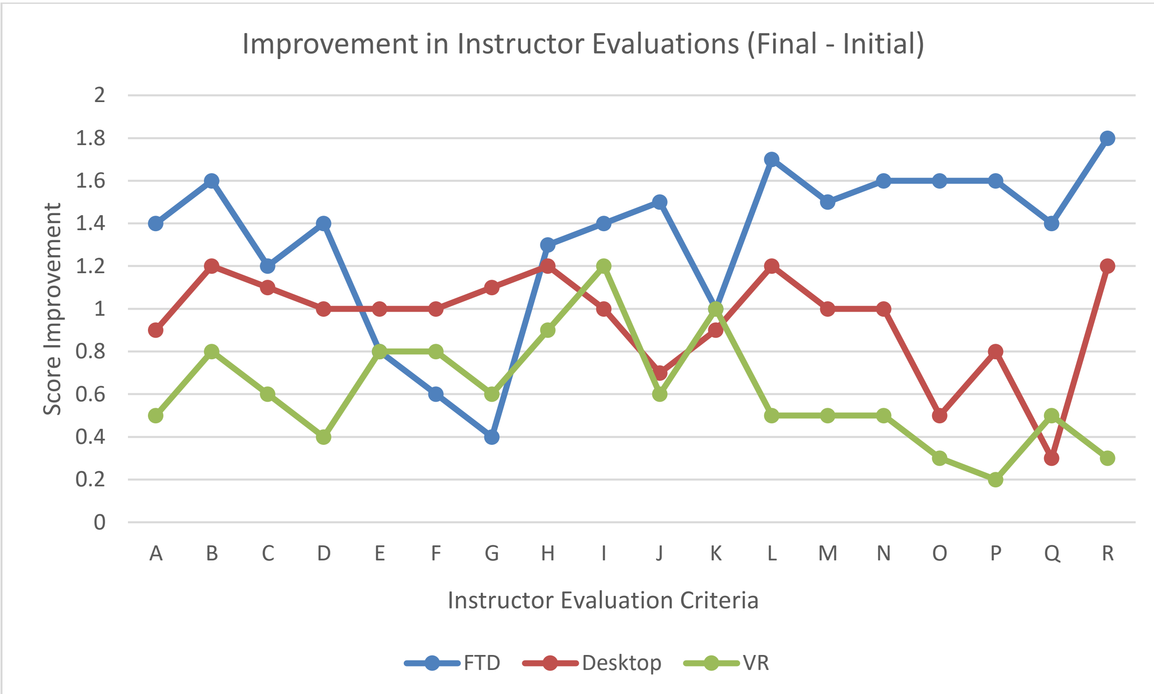


Figure 16: Improvement in Instructor Evaluations for FTD, Desktop, and Virtual Reality Groups

4.1.2 Flight Variability

Along with instructor evaluations, performance was also evaluated using flight variability of four key aircraft telemetry measures: altitude, heading, airspeed, and angle of bank.

For each measure, the standard deviation from each participant's flight was calculated. The improvement score was then calculated as the *Final Standard Deviation – Initial Standard Deviation* and compared using a one-way ANOVA to determine if there are any differences resulting from the type of simulator used for training. The Tukey pairwise comparison was also used to further investigate the results. As such, a positive score indicates an increase in standard deviation and thus worse performance, while a negative score indicates a decrease in standard deviation and thus improved performance.

First, considering the takeoff, the goal was for pilots to maintain a consistent speed (70-80 KIAS), heading (245°), and angle of bank (0°) during the first 1000 ft of the climb. As the average takeoff speed was reached after 11 seconds (SD = 8.5) and the 1000 ft of altitude gained by 89 seconds (SD = 13.24), the flight variability for the takeoff was analyzed from 10-90 seconds into the flight.

The mean improvement scores for the standard deviation of each aircraft telemetry measure during takeoff are included in Table 6. The more negative the score is, the better the improvement as smaller variation describes more stable aircraft control.

Table 6: Improvement in Standard Deviation of Aircraft Telemetry during Takeoff

		N	Mean	Std. Deviation
Heading	FTD	10	-33.366038033595700	4.888673885983915
	Desktop	10	-.792016292538279	1.791034870287563
	VR	10	34.827330520725950	4.935908180927131
	Total	30	.223092064863990	28.605741766990388
Airspeed	FTD	10	-.464185167793109	.584704489472589
	Desktop	10	-.735223098889773	1.117708339199392
	VR	10	.346247086059851	.945935661520988
	Total	30	-.284387060207677	.994923478644882
Bank	FTD	10	.655539350731259	1.419621623149737
	Desktop	10	.214548762002851	.612014069345501
	VR	10	-.237982420041551	.946174220744536
	Total	30	.210701897564186	1.075724609349582

From ANOVA, it was found that, first, the improvement in maintaining heading between groups training under different simulator conditions exhibited significant differences ($F(2,27) = 678.077$, $p < .001$, $\eta_p^2 = .980$). A pairwise comparison confirmed there was a significant difference between the FTD and desktop groups ($p <.001$), the desktop and VR groups ($p <.001$), and the VR and FTD groups ($p <.001$), with the VR group exhibiting the least improvement and the FTD group exhibiting the highest level of improvement.

The ANOVA also identified significant differences in the improvement of variance for airspeed between groups training under different simulator conditions ($F(2,27) = 3.821$, $p = .035$, $\eta_p^2 = .221$), with a significant difference specifically occurring between the desktop and VR groups ($p = .034$); the desktop groups exhibiting the highest level of improvement and the VR group exhibiting the least improvement. However, the differences between the FTD and desktop groups (p = .785) and the VR and FTD groups (p = .134) are non-significant.

Finally, there was no significant difference between simulator conditions for the variance of bank angle during takeoff ($F(2,27) = 1.823$, $p = .181$, $\eta_p^2 = .119$), whether that is between the FTD and desktop groups (p = .619), the desktop and VR groups (p = .604), or the VR and FTD groups (p = .155).

These results may also be visualized in Figure 17, Figure 18, and Figure 19.

Using a one-sample T test, it was found that, as seen in Figure 17, for the improvement in stability of heading, the improvement of the FTD group and the VR group are significantly different than 0 with $p < .001$. However, the desktop group did not demonstrate improvement different from 0 (p = .195).

Additionally, for improvement in stability of airspeed, as seen in Figure 18, the FTD group demonstrated improvement significantly different from 0 (p = .033), however, the desktop group (p = .067) and the VR group (p = .277) showed improvement similar to 0.

Finally, for improvement in stability of the angle of bank, as seen in Figure 19, all three groups did not demonstrate improvement different from 0. The FTD group, desktop group, and VR group had significance levels of p = .178, .296, and .447, respectively.

As there were significant differences exhibited between groups, with the VR group demonstrating worse improvement than participants who were a part of the higher-fidelity desktop and FTD groups, it may be concluded that, for aircraft handling tasks such as the takeoff, VR is insufficient to allow for proper training.

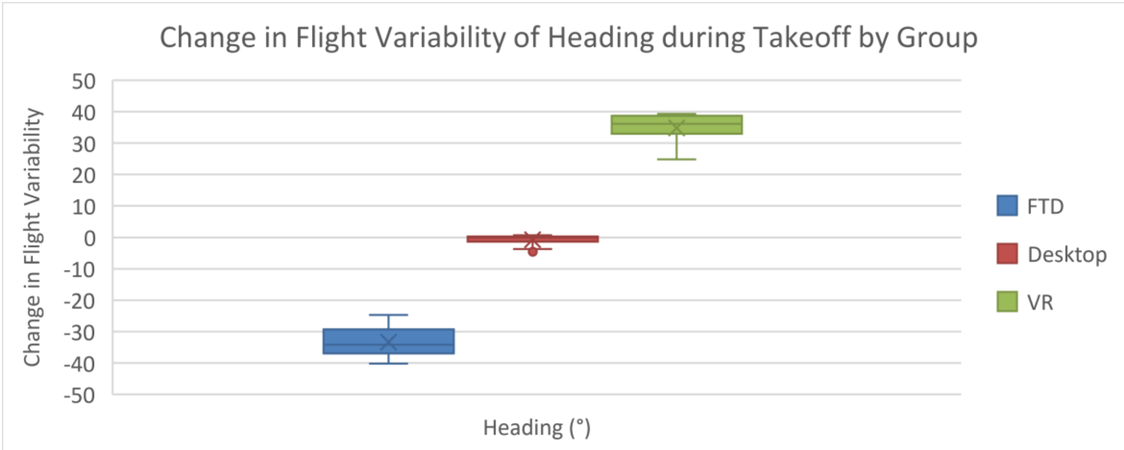


Figure 17: Change in Flight Variability of Heading during Takeoff by Group (Note: The more negative the score, the larger the improvement in aircraft control and flight stability)

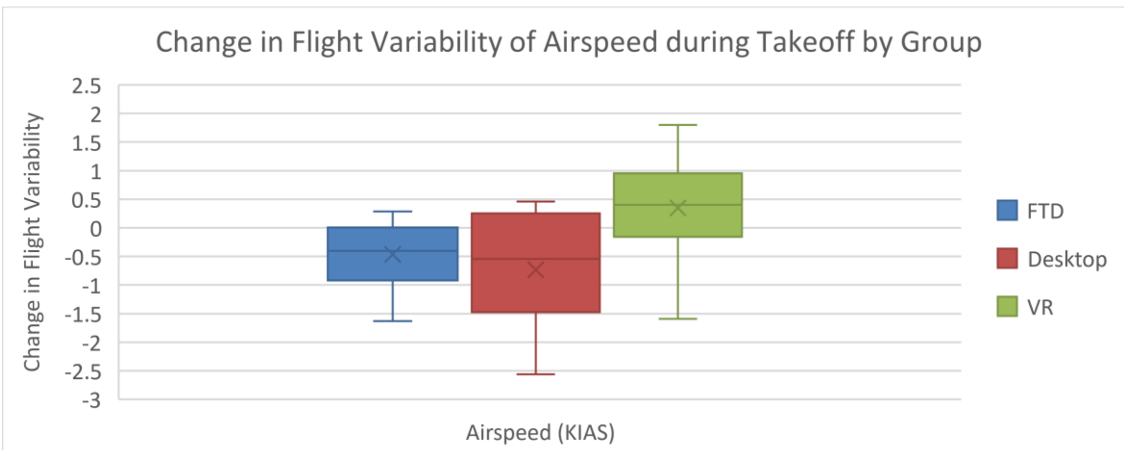


Figure 18: Change in Flight Variability of Airspeed during Takeoff by Group

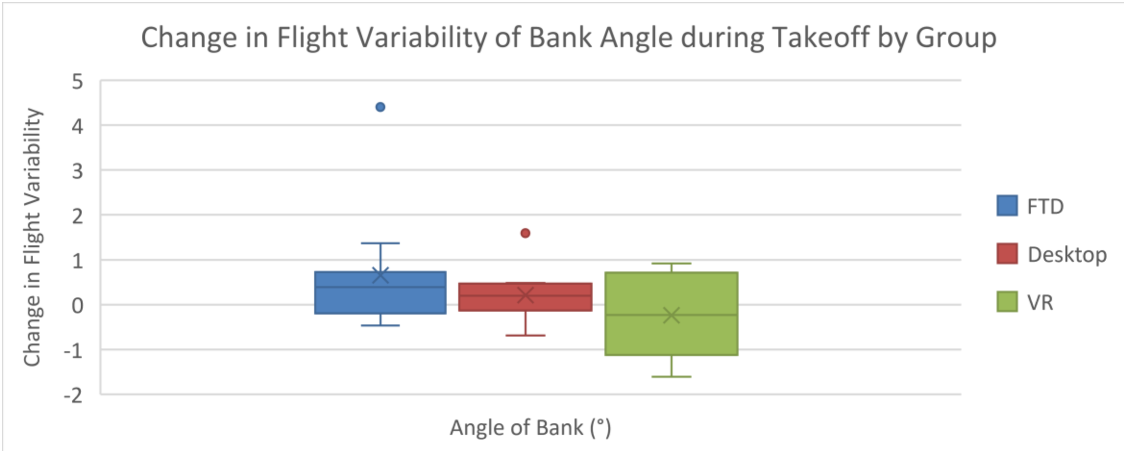


Figure 19: Change in Flight Variability of Bank Angle during Takeoff by Group

Second, considering the steep turn, the goal was for pilots to maintain a consistent altitude (4500 ft), airspeed (100 KIAS), and angle of bank (45°) while making a full 360° turn from North. Participants took on average 74 seconds (SD = 7.69) to complete the turn, with the simulation stopped upon completion of the turn. As such, the flight variability for the steep turn was analyzed for the final 80 seconds of each flight.

The mean improvement scores for the standard deviation of each aircraft telemetry measure during the steep turn are included in Table 7. Once again, the more negative the score, the larger the improvement as less variation means better aircraft control and greater stability.

Table 7: Improvement in Standard Deviation of Aircraft Telemetry during the Steep Turn

		N	Mean	Std. Deviation
Altitude	FTD	10	-16.778209346383495	10.999976467283878
	Desktop	10	-16.261966941590600	11.058740534404192
	VR	10	-7.610964746154283	12.015651112493074
	Total	30	-13.550380344709458	11.773029172786249
Airspeed	FTD	10	-1.490322111290318	1.054392489862171
	Desktop	10	-1.150896713488293	.822417160200069
	VR	10	-.510565264332539	.817457478896493
	Total	30	-1.050594696370384	.965931334680823
Bank	FTD	10	-.069578567315519	1.665008466604348
	Desktop	10	.518309048199444	1.402541570036432

	VR	10	-.249684268094021	.800829395938944
	Total	30	.066348737596635	1.334586931977736

From ANOVA, it was found that there was no significant difference in the improvement of aircraft telemetry measures between groups training under different simulator conditions for the steep turn for altitude ($F(2,27) = 2.053$, $p = .148$, $\eta_p^2 = .132$), airspeed ($F(2.27) = 3.023$, $p = .065$, $\eta_p^2 = .183$), or angle of bank ($F(2,27) = .899$, $p = .419$, $\eta_p^2 = .062$).

Furthermore, the pairwise comparison showed no significant difference between the FTD and desktop groups, desktop and VR groups, or VR and FTD groups for any of these three aircraft telemetry measures.

These results may also be visualized in Figure 20, Figure 21, and Figure 22.

Using a one-sample T test, it was found that, as seen in Figure 20, for the improvement in stability of altitude, the improvement of the FTD group and desktop group are significantly different than 0 with $p < .001$ and $p = .001$, respectively. However, the VR group did not demonstrate improvement different from 0 ($p = .076$).

Additionally, for improvement in stability of airspeed, as seen in Figure 21, the improvement of the FTD group and desktop group are significantly different from 0 with $p = .002$ for both groups. The VR group did not show improvement significantly different from 0 ($p = .080$).

Finally, for improvement in stability of the angle of bank, as seen in Figure 22, all three groups did not demonstrate improvement different from 0. The FTD group, desktop group, and VR group had significance levels of $p = .898$, .273, and .350, respectively.

As there were no significant differences exhibited between groups for the steep turn, it may be concluded that VR has potential for training some aircraft handling tasks in terms of improving stability in flight (i.e. reducing flight variability).

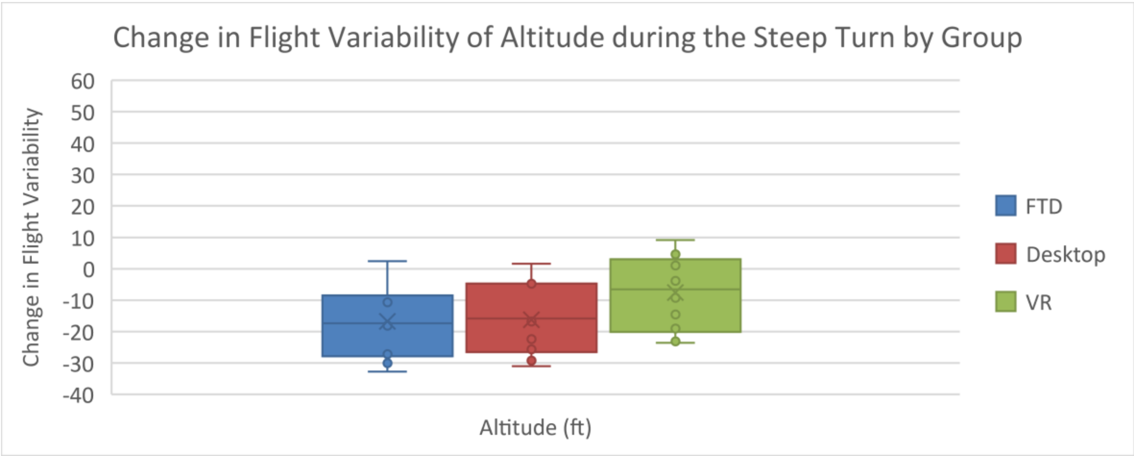


Figure 20: Change in Flight Variability of Altitude during the Steep Turn by Group

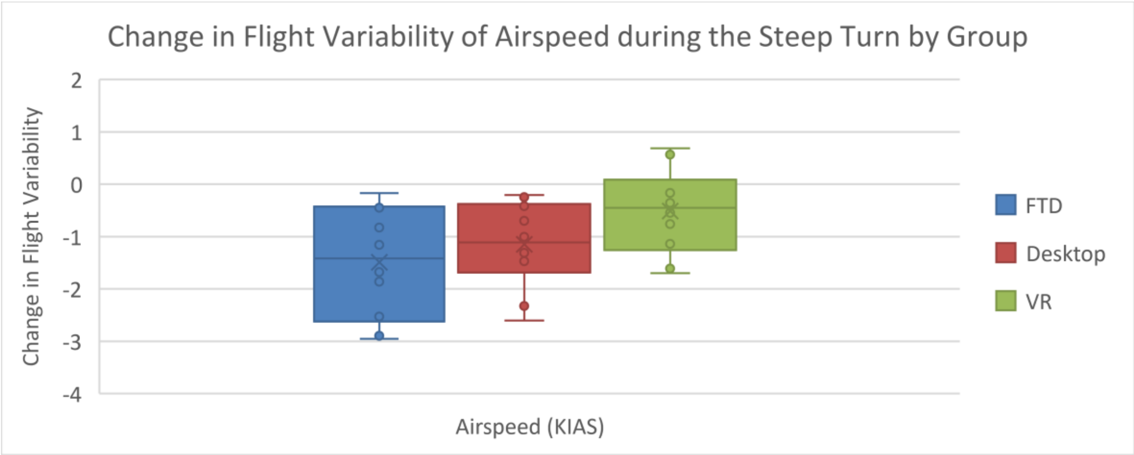


Figure 21: Change in Flight Variability of Airspeed during the Steep Turn by Group

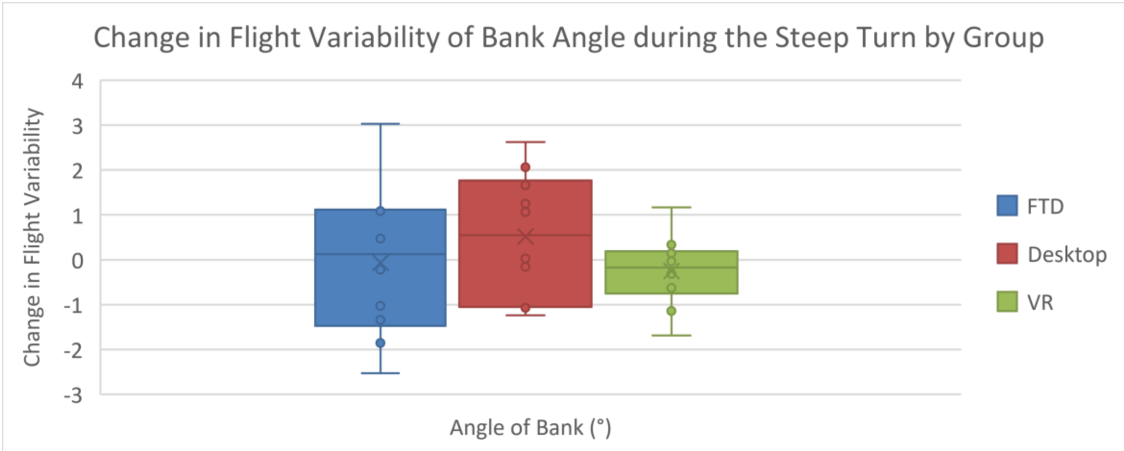


Figure 22: Change in Flight Variability of Bank Angle during the Steep Turn by Group

4.2 Cognitive Fidelity

While the real aircraft is always going to be a more stressful environment as the consequences of a mistake are quite severe in real flight settings, a proper training setting (including both the device and scenario used) can lead to comparable levels of mental workload for the same task. Thus, in evaluating the cognitive fidelity of the various simulators – the low-fidelity virtual reality simulator, medium-fidelity desktop simulator, and high-fidelity flight training device – the ideal is for the VR simulator to accurately replicate the stress and mental workload experienced in the higher-fidelity FTD. To accurately evaluate the cognitive fidelity, only scores obtained on day 3 are required, as this is the day where all participants were using their assigned simulator, and thus there is data available to compare the stress and mental workload between simulators.

4.2.1 Stress

Flight may be an inherently stressful situation, especially depending on the specific flight task; the closer that a simulator can replicate the stress of a flight, the better the student's learning. As such, the stress level of participants using the virtual reality simulator was evaluated and compared to that of participants who trained using the desktop simulator and FTD.

The closer the stress level of participants using the VR simulator matches that of participants using the FTD, the higher the cognitive fidelity of the VR simulator, and thus the more effective the VR simulator is for pilot training.

To conduct this evaluation, participants responded to the subjective stress scale, with the stress level caused by the simulation defined by *stress level pre-experiment (Day 3) – stress level post-experiment (Day 3)* using the questionnaires completed on day 3 of the experiment, when all participants were training using their assigned simulator. As such, a positive value indicates an increase in stress due to the simulation, a negative value indicates a decrease in stress due to the simulation, and a value of 0 indicates that the simulation had no effect on the stress level of the participant.

Table 8 shows the descriptive statistics for the difference in stress level due to simulator training for the VR, desktop, and FTD groups.

Table 8: Descriptive Statistics of the Difference in Stress Scores

Group	Mean	N	Std. Deviation
FTD	1.00	10	1.491
Desktop	.40	10	1.174
VR	.20	10	1.229
Total	.53	30	1.306

A between-subjects, one-way ANOVA was conducted to test for a difference in the change in stress level of participants between the VR, desktop, and FTD groups. There is no significant effect of the simulator type on stress level due to simulator training ($F(2.27) = .171$, $p = .844$, $\eta_p^2 = .012$). This result is also visualized in Figure 23, where it may be seen that the change in stress level of each group (VR, desktop, and FTD) have a similar mean and overlap in their interquartile range.

Further analyzing this result using a Tukey pairwise comparison, it may be seen that the change in stress was non-significant between the FTD and desktop groups ($p = .989$), the desktop and VR groups ($p = .842$), and the VR and FTD groups ($p = .907$).

Thus, it may be concluded that the cognitive fidelity of virtual reality is similar to that of the higher-fidelity simulators in terms of stress, and as such, virtual reality may be deemed to be an effective simulator for pilot training when considering this aspect of cognitive fidelity.

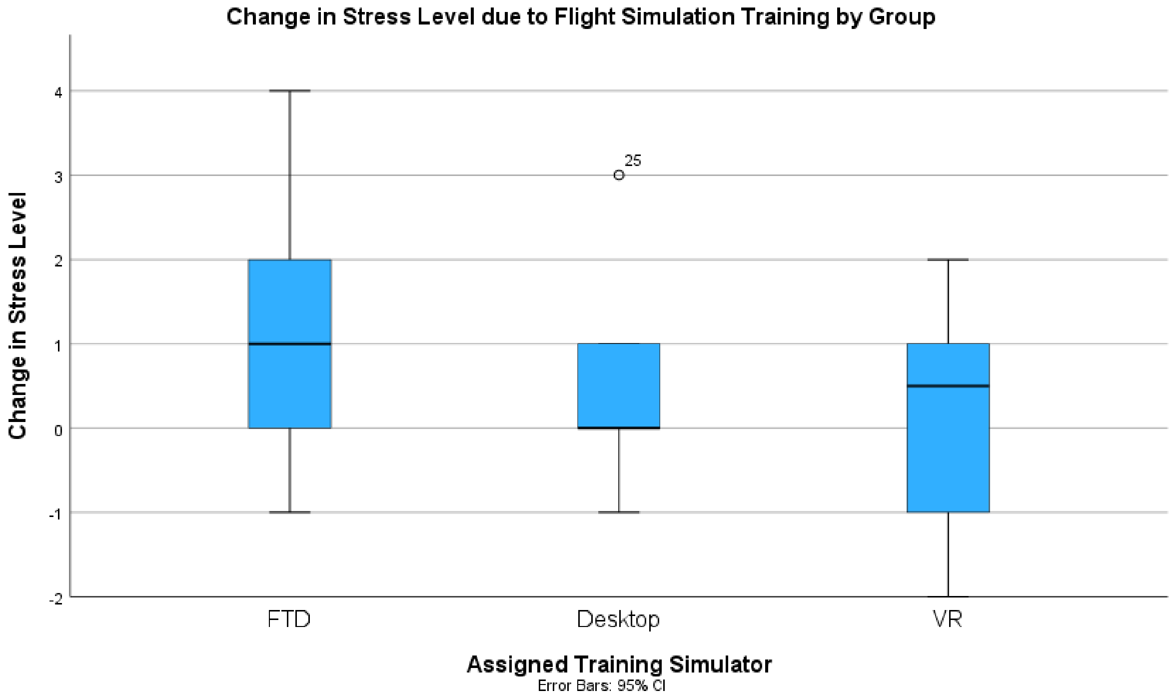


Figure 23: Difference in Stress Scores for FTD, Desktop, and VR Training Simulators

4.2.2 Mental Workload

Mental workload is another key aspect of cognitive fidelity which must be considered in an analysis of various simulators. Similar to stress, the closer the mental workload of participants using the virtual reality similar is to that of participants using the higher-fidelity flight training device, the higher the cognitive fidelity of the VR simulator and thus the more effective the VR simulator is for pilot training.

To evaluate mental workload, the NASA-TLX was given to participants after the experiment on Day 3 of the study. To keep the questionnaire to a reasonable length, pairwise comparisons between the sources of loads were not included; it is assumed that all sources of loads contribute equally to the overall workload.

For each of these six factors, participants were asked to rank their experience on a scale divided into 10 equal intervals anchored by the bipolar descriptors low/high (or good/poor in the case of performance). As such, each interval divided the scale from 0 to 100 in increments of 10. A lower score indicates low demand, effort, or frustration, and good performance, while a higher score indicates high demand, effort, or frustration, and poor performance.

Mental workload was evaluated for both procedural tasks (before start checklist) and aircraft handling tasks (takeoff and steep turn).

First, a repeated measures ANOVA was conducted to investigate the effect of the between subject's factor, type of simulator, and the within subject's factor, type of flight task (procedural tasks versus aircraft handling tasks).

From the repeated measures ANOVA, it was found that there is a significant effect of flight task type on mental workload ($F(1,27) = 48.182$, $p < .001$, $\eta_p^2 = .641$), with the mental workload being higher for aircraft handling tasks ($\mu = 4.56$, SD = 1.46) than procedural tasks ($\mu = 3.09$, SD = 1.33). There is no significant interaction between the type of flight task and the type of simulator used ($F(2,27) = .602$, $p = .555$, $\eta_p^2 = .043$).

Additionally, it was found that there is a significant effect of the type of simulator used on mental workload ($F(1,27) = 264.696$, $p < .001$, $\eta_p^2 = .907$).

For procedural tasks, the mental workload was higher for VR ($\mu = 3.33$, SD = 1.33) and desktop ($\mu = 3.43$, SD = 1.44) groups than it was for the FTD group ($\mu = 2.50$, SD = 1.30). Similarly, for aircraft handling tasks, the mental workload was higher for VR ($\mu = 4.40$, SD = 1.44) and desktop ($\mu = 5.00$, SD = 1.52) groups than it was for the FTD group ($\mu = 3.97$, SD = 1.38).

As significant effects were found, a one-way ANOVA was used to further analyze the impacts of the type of simulator on mental workload for both procedural and aircraft handling tasks.

4.2.2.1 Mental Workload in Procedural Tasks

Table 9 shows the descriptive statistics for the mental workload experienced by participants during completion of the procedural tasks, specifically the before start checklist.

Table 9: Descriptive Statistics for Mental Workload of Procedural Tasks

Group		Mental Demand	Physical Demand	Temporal Demand	Effort	Frustration	Performance
FTD	Mean	3.50	2.10	2.50	3.20	1.60	2.10
	N	10	10	10	10	10	10
	Std. Deviation	1.958	1.663	1.080	1.751	1.265	1.595
Desktop	Mean	4.20	2.80	2.80	4.70	1.80	4.30
	N	10	10	10	10	10	10
	Std. Deviation	2.781	1.687	1.476	2.669	.919	3.093
VR	Mean	3.60	2.60	3.00	4.00	2.40	4.40
	N	10	10	10	10	10	10
	Std. Deviation	1.955	1.075	1.944	1.826	.966	2.547
Total	Mean	3.77	2.50	2.77	3.97	1.93	3.60
	N	30	30	30	30	30	30
	Std. Deviation	2.208	1.480	1.501	2.141	1.081	2.634

A between-subjects, one-way ANOVA was conducted to test for a difference in the mental workload of participants between groups. There is no significant difference in the six factors contributing towards mental workload for procedural tasks between groups ($F(2,27) = .267$ to 2.726, $p = .084$ to .768, $\eta_p^2 =$ .146 to .372).

Further analyzing this result using a Tukey pairwise comparison, it was confirmed that there is not a significant difference in any of the six sources of mental workload between the FTD and desktop groups ($p = .138$ to .907), between the desktop and VR groups ($p = .427$ to .996), or between the VR and FTD groups ($p = .116$ to .995).

As participants using the VR, desktop, and FTD simulators experienced similar levels of mental workload for all six factors, we can conclude that the cognitive fidelity of VR is similar to that of the higher-fidelity simulators. As a result, for procedural tasks, in terms of mental workload and cognitive fidelity, VR may be an effective simulator method.

This supports the first hypothesis, that VR simulators are effective for training ab initio students in procedural tasks.

4.2.2.2 Mental Workload in Aircraft Handling Tasks

Table 10 shows the descriptive statistics for the mental workload experienced by participants during completion of the aircraft handling tasks, specifically take-off and the steep turn.

Table 10: Descriptive Statistics for Mental Workload of Aircraft Handling Tasks

Group		Mental Demand	Physical Demand	Temporal Demand	Effort	Frustration	Performance
FTD	Mean	4.70	4.30	3.80	4.60	3.00	3.40
	N	10	10	10	10	10	10
	Std. Deviation	2.163	1.567	1.549	1.776	1.333	1.174
Desktop	Mean	6.30	5.10	4.00	5.40	3.90	5.30
	N	10	10	10	10	10	10
	Std. Deviation	1.767	1.912	2.404	1.647	2.283	2.058
VR	Mean	5.10	3.40	3.50	5.50	3.90	5.00
	N	10	10	10	10	10	10
	Std. Deviation	2.470	1.776	1.900	1.780	2.234	1.826
Total	Mean	5.37	4.27	3.77	5.17	3.60	4.57
	N	30	30	30	30	30	30
	Std. Deviation	2.189	1.837	1.924	1.724	1.976	1.870

A between-subjects, one-way ANOVA was conducted to test for a difference in the mental workload of participants between groups. There is a significant difference in the performance of participants between groups ($F(2,27) = 3.499$, $p = .045$, $\eta_p^2 = .410$), however, the remaining five factors contributing towards mental workload do not show a significant difference in the mental workload for aircraft handling tasks between groups ($F(2,27) = .161$ to 2.342, $p = .115$ to $.852$, $\eta_p^2 = .114$ to $.350$).

Further analyzing this result using a Tukey pairwise comparison, it was confirmed that there was no significant difference in the mental demand, physical demand, temporal demand, effort, or frustration experienced by participations between the FTD and desktop groups ($p = .238$ to $.972$), between the desktop and VR groups ($p = .096$ to 1.000), or between the VR and FTD groups ($p = .487$ to $.939$).

Additionally, the difference in the performance experienced by participants between the FTD and desktop groups is on the cusp of being significant ($p = .052$), while there are non-significant results between the desktop and VR groups ($p = .920$) and between the VR and FTD groups ($p = .115$).

As differences between the desktop or VR groups and the FTD group being on the cusp of being significant, and the ANOVA showing significant effects of the type of simulator used on performance, it may be inferred that VR may not be an effective simulator for training aircraft handling tasks.

This result directly supports hypothesis 2 as it demonstrates that VR simulators have limitations in training aircraft handling tasks, leaving higher-fidelity FTDs as the ideal for training of these tasks.

4.3 Simulator Sickness

As virtual reality and other virtual environments created by simulators have been known to cause symptoms of simulator sickness, participants experience of simulator sickness during the study was evaluated using the Simulator Sickness Questionnaire (SSQ).

With high-fidelity flight training devices having been approved for use in pilot training, it is assumed that the level of simulator sickness experienced using these simulators has been deemed to be a safe level. As such, VR simulation will similarly be deemed safe for use in ab initio pilot training if the experience of simulator sickness symptoms in the virtual reality training group does not exceed that of the symptoms experienced by those in the FTD group.

To conduct this evaluation, participants responded to the SSQ on Day 3 of the study, which asked participants to rank their experience of 16 symptoms as “None”, “Slight”, “Moderate”, or “Severe”, both prior to and after the experiment, with participants using their assigned simulator for training. For each symptom, the labels were assigned a numerical value with 0 being “None” and 4 being “Severe”. The experience of simulator sickness caused by the simulation was then defined as *symptom level post-experiment (Day 3) – symptom level pre-experiment (Day 3)* for each of the 16 symptoms. As such, a positive value indicates the experience of simulator sickness during the experiment, a negative or ‘0’ value indicates that the simulation did not cause the experience of simulator sickness.

Table 11 shows the descriptive statistics for the experience of simulator sickness due to simulator training using the VR, desktop, or FTD simulators.

Table 11: Descriptive Statistics for the Experience of Simulator Sickness

Group		General Discomfort	Fatigue	Headache	Eye Strain	Difficulty Focusing	Increased Salivation
FTD	Mean	.10	-.10	-.20	-.10	.00	-.10
	N	10	10	10	10	10	10
	Std. Deviation	.738	.316	.632	.316	.000	.316
Desktop	Mean	.10	-.60	.00	.10	.00	-.20
	N	10	10	10	10	10	10
	Std. Deviation	.316	1.265	.471	.568	.000	.422
VR	Mean	.40	.00	.20	.30	.00	.10
	N	10	10	10	10	10	10
	Std. Deviation	.699	.000	.789	.675	.000	.316
Total	Mean	.20	-.23	.00	.10	.00	-.07
	N	30	30	30	30	30	30
	Std. Deviation	.610	.774	.643	.548	.000	.365

Group		Sweating	Nausea	Difficulty Concentrating	Fullness of Head	Blurred Vision
FTD	Mean	.30	.00	-.30	-.10	.10
	N	10	10	10	10	10
	Std. Deviation	.675	.000	.483	.316	.316
Desktop	Mean	.00	.00	-.10	.00	-.10
	N	10	10	10	10	10
	Std. Deviation	.471	.000	.568	.000	1.101
VR	Mean	.30	.20	.00	.20	.00
	N	10	10	10	10	10
	Std. Deviation	.675	.632	.000	.422	.000
Total	Mean	.20	.07	-.13	.03	.00
	N	30	30	30	30	30
	Std. Deviation	.610	.365	.434	.320	.643

Group		Dizzy Eyes Open	Dizzy Eyes Closed	Vertigo	Stomach Awareness	Burping
FTD	Mean	.00	.00	.00	.00	.10
	N	10	10	10	10	10
	Std. Deviation	.000	.000	.000	.000	.316
Desktop	Mean	.00	.00	.00	.00	-.10
	N	10	10	10	10	10
	Std. Deviation	.000	.000	.000	.000	.316
VR	Mean	.50	.20	.00	.20	.00
	N	10	10	10	10	10
	Std. Deviation	.972	.632	.000	.632	.000
Total	Mean	.17	.07	.00	.07	.00
	N	30	30	30	30	30
	Std. Deviation	.592	.365	.000	.365	.263

While there are standards for calculating symptoms for overall nausea, oculomotor disturbance, and disorientation, as well as overall total simulator sickness (Kennedy et al., 1993), there is no standard method of evaluating simulator sickness scores which, on their own, may indicate significant or concerning amounts of simulator sickness, or in interpreting symptoms which have been reduced after exposure to the experiment (Pauline Bimberg et al., 2020).

As such, for the analysis of data for this study, simulator sickness scores were analyzed as a difference between the pre- and post-experiment questionnaires for each of the 16 symptoms individually.

A between-subjects, one-way ANOVA was conducted to test whether the experience of simulator sickness in the VR group was significantly different than that of the simulator sickness experienced in the desktop and FTD groups. There are no significant effects of type of simulator used on simulator sickness ($F(2,27) = .229$ to 2.647, $p = .089$ to $.797$, $\eta_p^2 = .017$ to $.164$). These results are visualized in Figure 24.

Further analyzing these results using a Tukey pairwise comparison, it was found that there are no significant differences in the experience of simulator sickness for all 16 symptoms between the FTD and desktop groups ($p = .212$ to 1.00), between the desktop and VR groups ($p = .133$ to $.939$), and between the VR and FTD groups ($p = .088$ to 1.00).

As there is no difference in the experience of simulator sickness between simulators for any of the 16 simulator sickness symptoms, it is concluded that VR simulators are a safe alternative to higher-fidelity simulators, such as the already approved FTDs.

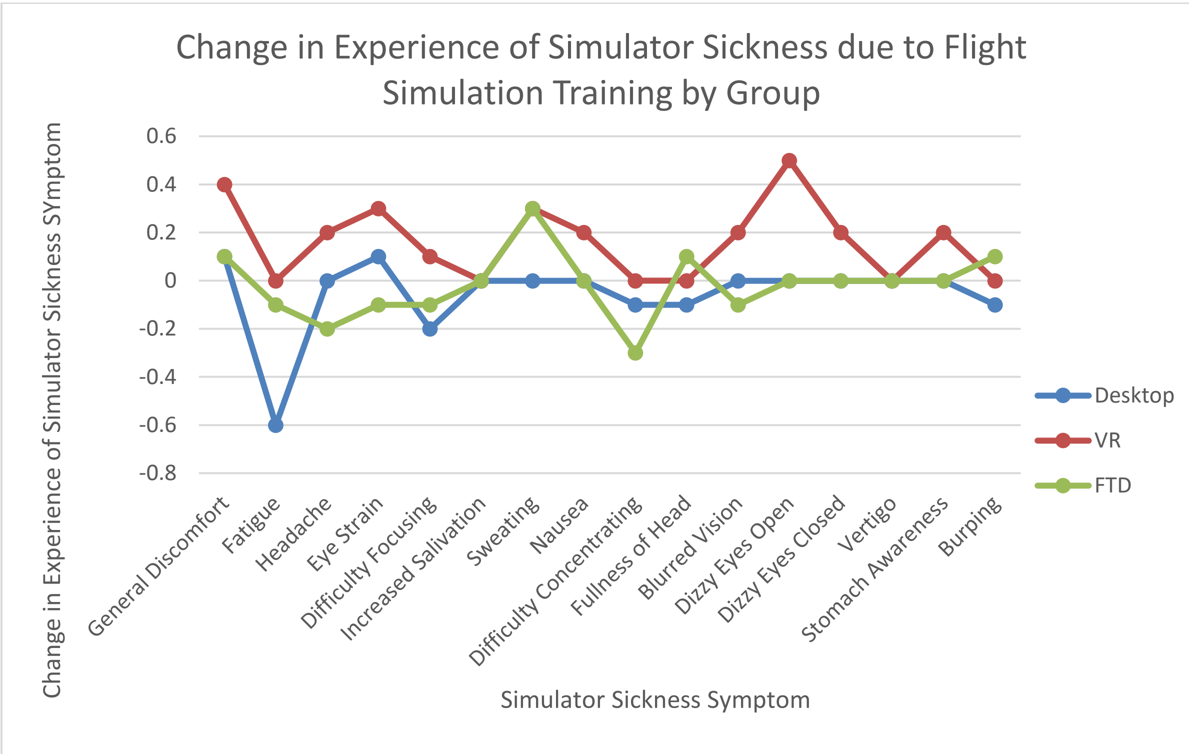


Figure 24: Change in Experience of Simulator Sickness by Group

4.4 Summary of Results

4.4.1 Hypothesis 1: Procedural Tasks

Recall that the first hypothesis predicted that virtual reality head-mounted displays would be as effective as traditional high-fidelity simulators for training procedural tasks to ab initio student pilots.

From the instructor evaluations, it was determined that the VR group demonstrated an improvement in performance which is not significantly different to that of the higher-fidelity desktop and FTD simulators for half the criteria of the before start checklist, as well as for the procedural components of the takeoff. It is predicted that the lack of similarity of the FTD to the actual Cessna 172 cockpit configuration was the cause for criteria in which VR did not lead to similar performance for procedural criteria, and thus, overall, VR may still be deemed to be effective at training procedural tasks in terms of performance.

Additionally, it was determined that, for procedural tasks, the VR simulator effectively replicated the cognitive fidelity of the higher FTD in terms of both stress and mental workload, as again there were no statistically significant differences found in stress or mental workload, again providing evidence towards the potential use of VR in training procedural flight tasks.

In summary, this research has shown that VR may be effectively used to train procedural tasks for ab initio student pilots with performance results similar to that of the currently used higher-fidelity flight training devices.

4.4.2 Hypothesis 2: Aircraft Handling Tasks

Recall that the second hypothesis predicted that virtual reality head-mounted displays would not be sufficient for training aircraft handling tasks to ab initio student pilots, and thus traditional high-fidelity simulators would still be required for ab initio pilot training.

From the instructor evaluations, for the aircraft handling components of the takeoff as well as for the steep turn, all three groups demonstrated improvement from Day 1 to Day 5, however, participant's improvement varied significantly as a result of the type of simulator used for training, with the VR group demonstrating significantly less improvement than the high-fidelity FTD group. This result was confirmed through analysis of flight variability, which demonstrated that there were significant differences in improvement in terms of the aircraft telemetry measures of heading and airspeed. Despite the VR group demonstrating similar performance in terms of aircraft telemetry measures for the steep turn, overall, there is evidence to suggest that VR is not as effective at training aircraft handling tasks, and thus, traditional high-fidelity simulators are still required for training these specific tasks.

Additionally, although it was determined that the VR simulator replicated the cognitive fidelity of the higher FTD in terms of showing no statistically significant differences in stress, the mental workload of participants using the VR simulator for training was significantly higher than that of participants using the FTD, specifically considering performance.

In conclusion, this research has demonstrated that, while VR is able to train aircraft handling tasks, it is unable to train these tasks for ab initio student pilots to the same degree of success as the currently used higher-fidelity flight training devices.

4.4.3 Simulator Sickness

An analysis into the occurrence of simulator sickness (sometimes referred to by other terminology, including Simulator Adaptation Syndrome (SAS)) of participants found that there was no greater risk of experiencing symptoms of simulator sickness when using the virtual reality simulator or desktop simulator compared to that of participants using the high-fidelity flight training device.

Additionally, as dropouts are typically indicative of severity of simulator sickness, it is important to note that no participants dropped out of the study at any stage.

As such, concern around the potential occurrence of simulator sickness should not be a major factor determining the use lower-fidelity simulators including virtual reality simulators or desktop simulators, in ab initio flight training.

Chapter 5
Discussion

The main goal of this study was to investigate the potential for use of virtual reality HMDs in ab initio pilot training. The main research questions in this study were to address gaps in the existing literature through an experimental comparison between high-fidelity flight training devices (FTDs), medium-fidelity desktop simulators, and low-fidelity virtual reality (VR) simulators, specifically considering (1) differences in improvement in performance for ab initio procedural pilot training tasks, (2) differences in improvement in performance for ab initio aircraft handling pilot training tasks, (3) differences in cognitive fidelity including mental workload and stress, and (4) the occurrence of simulator sickness. This chapter discusses the outcomes of this study with a focus on the original research questions, using the results reported in Chapter 4 as evidence. As there exists a very limited amount of literature covering research questions similar to this study, outcomes were mainly evaluated based on a comparison to the simulators currently approved for training.

One of the major concerns with implementing virtual reality training in place of existing training is the lack of natural tactile interaction, which previous research has implied is essential for successful training. However, in discussion with current pilots, it was identified that natural tactile interaction, that the feel of and feedback from flight controls and instrument panels, is not necessary in all aspects of pilot training, specifically procedural tasks. Currently, procedure trainers are used to ensure pilots are familiar with the cockpit configuration and the specific procedures for the model of aircraft they are flying, before using the high-fidelity flight training devices. These procedure trainers may be as simple as a paper poster with a picture of the cockpit configuration, or may have a higher physical fidelity, modelling the cockpit, but with touchscreen tablets in place of physical instrument panels. VR has the most potential in pilot training as a higher-fidelity alternative to these procedure trainers, as has already been demonstrated in previous studies which have found VR to be significantly better at engagement, satisfaction, usability, presence, knowledge gain, retention, and confidence for procedural tasks, when compared to lower-fidelity devices such as a smartphone or paper (Buttussi & Chittaro, 2021). However, there is a lack of knowledge in the current literature comparing VR to the higher-fidelity simulators used in pilot training, which is the gap this study hoped to address.

First, considering differences in improvement in performance for ab initio procedural pilot training tasks, the instructor evaluations revealed no significant differences in improvement in performance for the before start checklist between groups training on any of the three types of simulators studied. This evidence shows that VR can be used in ab initio pilot training of procedural tasks.

Second, considering differences in improvement in performance for ab initio aircraft handling tasks, the instructor evaluations and flight variability data revealed significant differences, with participants training on the low-fidelity VR simulator improving significantly less that those training on the higher-fidelity simulators. As predicted due to the lack of natural tactile interaction in VR, this evidence shows that VR should not replace existing high-fidelity flight training devices for ab initio pilot training of aircraft handling tasks.

Third, considering cognitive fidelity, it was found that all three simulators led to similar levels of stress for the tasks studied. Additionally, for procedural tasks, the mental workload experienced by participants was similar for all types of simulators. However, for aircraft handling tasks, the VR group demonstrated significantly greater levels of mental workload. This evidence supports the two conclusions above: VR may effectively replace higher-fidelity simulators for procedural training tasks but should not be used as a replacement when training aircraft handling tasks.

Fourth and finally, considering the occurrence of simulator sickness, it was found that there were no significant differences in the severity of symptoms experienced between groups training on each type of simulator, and as such low-fidelity VR simulators may safely be used in pilot training alongside existing medium-fidelity desktop simulators and high-fidelity flight training devices. For the purpose of this study, time spent training during each session was limited to 30 minutes per the recommended guideline by Oculus that users take a break every 30 minutes while becoming accustomed to the headset (*Health and Safety Warnings,* n.d.). However, this duration is much shorter than what would be anticipated for pilot training, thus, further studies may be needed to properly analyze the occurrence of simulator sickness following a more realistic training regimen.

While further studies will need to be conducted to provide additional evidence supporting the use of VR HMDs in pilot training before regulators such as Transport Canada will approve them for use, some airlines have begun to implement VR training into their existing training protocols as they search for evidence supporting the use of VR which would ultimately lead to cost savings.

At Alaska Airlines, VR is being introduced in the classroom as a part of textbook training, that is training that occurs before ever entering a full flight simulator. As the Managing Director stated, “it’s helping give [the student pilots] exposure to things we typically wouldn’t see early in training which makes [the student pilots] more successful in the latter stages of training," (Kristin Goodwillie, 2022). Student pilots who have been some of the first to test the VR technology have made statements supporting the use of VR as a procedure training, stating "it's like your alarm clock in the morning, you know exactly where the snooze button is and this helps our muscle memory to know exactly where that switch is," and "when you start your first simulator sessions, we don't need to spend four hours trying to figure out where the switches are. You can step in on day one, minute one and know exactly where things are," (Kristin Goodwillie, 2022). This is helping improve the efficiency of training, allowing Alaska Airlines to increase their hiring pool and helping to address the ongoing pilot shortage. Alaska Airlines believes the technology is so successful that it will soon be in standard use throughout the aviation industry.

Along with improving the efficiency of training, VR has also been implemented at Embry-Riddle Aeronautical University for the first time in 2022, and was found to reduce the time it took student pilots to complete their first solo flight by over 30%, as well as prepared students to "immediately control the airplane and understand radio communications very well," (*New US University Flight Training Program Using Virtual Reality Cuts Time To Solo By 30%*, 2022).

With such successful results found in the industry, combined with scientific evidence supporting the use of VR, it seems likely that VR will become a standard part of pilot training implemented across the aviation industry in the future.

Current literature also points towards the use of VR throughout the aviation industry, not just focused on pilots, as Lufthansa Aviation Training has successfully implemented VR training for cabin crew upon the technology having been approved by the German Aviation Authority with hand tracking implemented (*VR Flight Training: Lufthansa Aviation Training & NMY*, 2022) and scientific literature strongly suggests the effectiveness of VR training for maintenance personnel (Alasim & Almalki, 2021; Brown et al., 2021; Jamain & Kasirun, 2011; Washburn et al., 2007; Yuviler-Gavish & Gopher, 2013).

As such, by decreasing the cost and time of training all personnel in the aviation industry, including pilots, cabin crew, and maintenance personnel, VR has the potential to significantly aid the industry in addressing the global personnel shortage, provided that scientific evidence supporting VR may be demonstrated, thus leading CAAs to approve VR for use in training.

Chapter 6
Conclusion

6.1 Summary

The present study was conducted with the main goal being to examine the effectiveness of virtual reality head-mounted displays (specifically looking at use of the VR HMD individually, without any added controls and instrumentation) in comparison to higher-fidelity desktop simulators and flight training devices which are the only flight simulators currently approved for use in pilot training. Participants trained to complete the before start checklist, takeoff, and a steep turn using either a VR simulator, desktop simulator, or FTD in ideal weather conditions. To successfully compare participants' performance between groups, all participants, regardless of the simulator assigned for their training, completed the flight tasks on the first and final day using the FTD.

Their performance was evaluated using both instructor evaluations recorded during the task by a full-time pilot, as well as flight variability data analyzed from the simulator output. Participants mental workload and stress level were also evaluated to allow for further comparisons in the cognitive fidelity of each simulator type. Additionally, participants were asked to complete the simulator sickness questionnaire to determine any adverse symptoms brought on by training using each type of simulator.

Analyzing performance as well as the simulators cognitive fidelity, it was determined that, for procedural tasks, VR led to similar increases in performance as the higher-fidelity alternatives, with comparable cognitive fidelity in terms of mental workload and stress experienced.

In contrast, it was determined that, for aircraft handling tasks, participants training with VR were unable to achieve the same increases in performance as the higher-fidelity alternatives, while also experiencing higher levels of mental workload.

There were no significant differences in terms of the experience of simulator sickness.

6.2 Limitations

The minimum sample size necessary, as computed using GPower, was the sample size of the study. However, given the small effect sizes obtained through the various analyses, a handful of effect sizes between 0.2 and 0.3, it is likely that the sample size was too small to obtain meaningful results for all the differences between groups. Using a post hoc power analysis for a repeated measures, between factors ANOVA on GPower 3.1.9.4 (Franz Faul et al., 2019), for an effect size of 0.3 the achieved power is solely 0.34, much lower than the desired 0.80. Using an a priori analysis with an effect size of 0.3, it is recommended to have a sample size of 84 (28 participants per group).

Additionally, in recording instructor evaluations, the mental workload on the instructor is quite high as they are responsible for reading the flight task checklists to participants, observing the instruments during the participant's flight, and recording their scores for the participant's performance. As such, there is a small probability of human error in the recorded values.

Aside from flight performance, cognitive fidelity of the simulators was also evaluated using the NASA-TLX and Subjective Stress Scale. While these measures have been proven to be reliable through their use in similar studies in the field of aviation, these measures are both subjective, which introduces variability and biases from each participant. If a study expanding upon the analysis of cognitive fidelity of these simulators were to be done in the future, it may be beneficial to use physiological or other objective measures in combination with these subjective measures to gain insight that does not rely on human judgement while still recording participants personal experiences.

Additionally, the tasks studied in this experiment were basic tasks taught in ab initio training, which are relatively stress-free compared to simulator training for emergency situations. If a study were to expand upon the analysis of cognitive fidelity of these simulators in the future, it may also be beneficial to study the cognitive fidelity of these simulators for higher-stress tasks.

As the study was conducted using three different types of flight simulators, the limitations of these simulators must also be considered. Due to cost and ethical considerations, the high-fidelity flight training device was used in place of the actual Cessna 172 aircraft. However, the FTD available for this study does not accurately replicate the cockpit layout, including the position of the instruments and switches, whereas the medium-fidelity desktop simulator and low-fidelity virtual reality simulator more accurately replicate the Cessna 172 cockpit configuration. Hence, it is possible that participants in the desktop and virtual reality groups would demonstrate better training transfer to the real Cessna 172 aircraft than was observed in transfer to the FTD.

Also, for the virtual reality simulator, the built-in VR controllers were used as input devices, which limits the pilot's ability to interact in a natural way with the flight controls and instruments. Specifically, this led to a lack of feedback from the yoke, as well as no interaction with rudder pedals, which led to improper flight conditions. Implementing proper yoke and rudder control to the VR simulator increases the fidelity of a VR simulator, which may lead to VR providing significantly better aircraft handling training than was observed in this study.

6.3 Implications

This study has found that, in terms of procedural flight tasks, there is no significant difference in the performance and experience of ab initio student pilots training with a low-fidelity virtual reality HMD, medium-fidelity desktop simulator, or high-fidelity flight training device. Additionally, participants using the VR simulator experienced similar levels of mental workload and stress as that of participants using the higher-fidelity simulators. There were also no significant differences in the overall occurrence of simulator sickness symptoms, implying that all three types of simulators are safe to use in this regard. Given this result, it is implied that VR may be effectively used in ab initio pilot training to prepare students to complete procedural tasks, such as the before start checklist, without compromising the quality of training or the level of safety. This would thus decrease each student's overall time spent learning procedural tasks in the higher-fidelity flight training devices, leading to savings in the overall cost of pilot training. Additionally, this would decrease the demand for use of flight training devices, making these high-fidelity devices more accessible.

Alternatively, for aircraft handling tasks, participants training on the VR simulator did not demonstrate improvement in performance up to the same level as those participants using the FTD or the desktop simulator. When using VR for aircraft handling tasks, the mental workload of participants was also significantly higher. Hence, VR is insufficient for training ab initio pilots in aircraft handling tasks, such as takeoff and steep turns. The use of high-fidelity flight training devices in training of aircraft handling tasks should continue to be the requirement in this case, ensuring the highest level of training and thus the highest standard for safety.

This study has provided data which allows flight training schools to more accurately conduct a cost benefit analysis comparing VR simulators, FTDs, and desktop simulators. This data may allow training schools to balance the use of these three types of simulators in a way which maximizes the effectiveness of training by utilizing the benefits of all three simulators, while simultaneously minimizing the overall training cost.

6.4 Future Work

Due to a lack of previous literature upon which to base the effect size, it was assumed in calculation of the sample size that a large effect size of 0.5 would be determined. However, the effect sizes found in the analysis were oftentimes much smaller, and as such, future work should include a larger sample size to accommodate for the smaller effect sizes.

Additionally, this study focused on comparing virtual reality and desktop simulators to a high-fidelity flight training device. However, this FTD did not accurately replicate the Cessna 172 cockpit configuration, possibly impacting the overall results. In order to replicate this data and the corresponding finding, one would be required to select an FTD that similarly used a different cockpit configuration than the actual Cessna 172 aircraft. It may be beneficial for future work to compare these lower-fidelity simulators to the real Cessna 172 aircraft, or a higher-fidelity flight training device that properly replicates the cockpit configuration of the desired aircraft.

Moreover, there are numerous flight tasks, including both procedural and aircraft handling tasks, which are taught to ab initio pilots. This study simply tested three tasks: the before start checklist, takeoff, and a steep turn. Future studies may expand upon this work by examining different procedural and aircraft handling tasks to ensure consistency in the results between these two flight task categories.

Finally, the major limitation of the virtual reality simulator is the lack of natural tactile interaction. However, it is possible to connect physical flight controls including rudder pedals and a yoke to the virtual reality HMD running X-Plane 11. In adding these controls, the natural tactile interaction of the virtual reality simulator would be improved to a point which may allow for effective training of aircraft handling tasks. Conducting a similar study under these conditions would be beneficial in understanding the full potential of virtual reality both with and without added flight controls.

References

172S Skyhawk: Information Manual. (2004). Cessna.

Agata Lawrynczyk. (2018). *Exploring Virtual Reality Flight Training as a Viable Alternative to Traditional Simulator Flight Training* [MASc, School of Computer Science, Carleton University]. https://curve.carleton.ca/system/files/etd/8befeba3-7102-4851-a690-7a513e0f8343/etd_pdf/ae488304b192921ff4725138d68a9976/lawrynczyk-exploringvirtualrealityflighttrainingasaviable.pdf

Airplane Pilot Demographics and Statistics in the US. (n.d.). Zippia. Retrieved October 21, 2022, from https://www.zippia.com/airplane-pilot-jobs/demographics/

AL250: The Right Device for your PPL, CPL, % IR/ME Training Needs. (n.d.). ALSIM Flight Training Solutions. Retrieved December 8, 2022, from https://www.alsim.com/simulators/al250/

Alasim, F., & Almalki, H. (2021). Virtual Simulation-Based Training for Aviation Maintenance Technicians: Recommendations of a Panel of Experts. *SAE International*. https://doi.org/10.4271/2021-01-0038

Alce, G., Klang, K.-J., Andersson, D., Nyström, S., Wallergård, M., & Niehorster, D. C. (2020). *Using Augmented Reality to Train Flow Patterns for Pilot Students—An Explorative Study*. *12242*, 215–231. https://doi.org/10.1007/978-3-030-58465-8_17

Aviation Personnel Licensing Statistics. (2022, March 29). Government of Canada | Gouvernement Du Canada. https://tc.canada.ca/en/aviation/licensing-pilots-personnel/aviation-personnel-licensing-statistics

Balk, S. A., Bertola, M. A., & Inman, V. W. (2013). *Simulator Sickness Questionnaire: Twenty Years Later*. Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Bolton Landing New York, United States. https://drivingassessment.uiowa.edu/sites/drivingassessment.uiowa.edu/files/wysiwyg_uploads/041_balk_0.pdf

Boas, Y. A. G. V. (n.d.). Overview of Virtual Reality Technologies. *Interactive Multimedia Conference*, *2013*. Retrieved June 7, 2021, from https://static1.squarespace.com/static/537bd8c9e4b0c89881877356/t/5383bc16e4b0bc0d91a758a6/1401142294892/yavb1g12_25879847_finalpaper.pdf

Brown, C., Hicks, J., Rinaudo, C. H., & Burch, R. (2021). The Use of Augmented Reality and Virtual Reality in Ergonomic Applications for Education, Aviation, and Maintenance. *Ergonomics in Design*. https://doi.org/10.1177/10648046211003469

Buttussi, F., & Chittaro, L. (2018). Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. *IEEE Transactions on Visualization and Computer Graphics*, *24*(2). https://doi.org/10.1109/TVCG.2017.2653117

Buttussi, F., & Chittaro, L. (2021). A Comparison of Procedural Safety Training in Three Conditions: Virtual Reality Headset, Smartphone, and Printed Materials. *IEEE Transactions on Learning Technologies*, *14*(1). https://doi.org/10.1109/TLT.2020.3033766

Chittaro, L., Buttussi, F., & Zangrando, N. (2014). Desktop Virtual Reality for Emergency Preparedness: User Evaluation of an Aircraft Ditching Experience Under Different Fear Arousal Conditions. *ACM Symposium on Virtual Reality Software and Technology*, 141–150. https://doi.org/10.1145/2671015.2671025

Chittaro, L., Corbett, C. L., McLean, G. A., & Zangrando, N. (2018). Safety Knowledge Transfer Through Mobile Virtual Reality: A Study of Aviation Life Preserver Donning. *Safety Science*, *102*, 159–168. https://doi.org/10.1016/j.ssci.2017.10.012

Christopher K. McClernon, Michael E. McCauley, Paul E. O'Connor, & Joel S. Warm. (2011). Stress Training Improves Performance During a Stressful Flight. *Human Factors*, *53*(3), 207–218. https://doi.org/10.1177/0018720811405317

Cockpit Procedure Trainers—Paper Tigers—Cardboard Bombers. (n.d.). Flight Vectors.

Control Loading. (n.d.). Redbird Flight Simulations, Inc. Retrieved May 9, 2022, from https://simulators.redbirdflight.com/technologies/control-loading

Dahai Liu, Nikolas D. Macchiarella, & Dennis A. Vincenzi. (2008). Chapter 4: Simulation Fidelity. In *Human Factors in Simulation and Training* (pp. 61–73). https://www.researchgate.net/profile/Dennis-Vincenzi-2/publication/278914678_Simulation_Fidelity/links/559fd91908aedb0e66f430d3/Simulation-Fidelity.pdf

Desktop Simulators. (n.d.). Redbird Flight Simulations, Inc. Retrieved May 9, 2022, from https://simulators.redbirdflight.com/products/topic/desktop-simulators

Drake R. Bradley & Stuart B. Abelson. (1995). Desktop flight simulators: Simulation fidelity and pilot performance. *Behavior Research Methods, Instruments, and Computers*, *27*(2), 152–159.

FAA Approval. (n.d.). Redbird Flight Simulations, Inc. Retrieved May 9, 2022, from https://simulators.redbirdflight.com/certification/faa

FAA Issues New Flight Simulator Regulations. (2016, April 14). Federal Aviation Administration. https://www.faa.gov/newsroom/faa-issues-new-flight-simulator-regulations

FAA-Certified X-Plane. (n.d.). X-Plane. Retrieved May 9, 2022, from https://www.x-plane.com/pro/certified/

Fees and Financing Your Education. (n.d.). University of Waterloo. Retrieved October 10, 2022, from https://uwaterloo.ca/aviation/fees

Franz Faul, Axel Buchner, Edgar Erdfelder, & Albert-Georg Lang. (2019). *GPower* (3.1.9.4) [Mac and Windows]. Behavior Research Methods. https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

Hannah Francis. (2014, June 10). *How to Deal with Stress as a Project Manager*. Wellingtone. https://wellingtone.co.uk/how-to-deal-with-stress-as-a-project-manager/

Health and Safety Warnings. (n.d.). Oculus.

Hoffman, H., & Vu, D. (1997). Virtual Reality: Teaching Tool of the Twenty-first Century? *Academic Medicine*, *72*(12), 1076–1081.

How much does a Frasca simulator cost? (n.d.). *Frasca Flight Simulation*. Retrieved May 19, 2021, from https://www.frasca.com/how-much-does-a-frasca-simulator-cost/

IBM SPSS (29.0.0.0 (241)). (n.d.). IBM. Retrieved January 11, 2023, from https://www.ibm.com/spss

In The Cockpit. (n.d.). The Basics of Aviation. Retrieved December 9, 2022, from https://sites.google.com/site/thebasicsofaviation/in-the-cockpit

Jamain, N. S., & Kasirun, Z. M. (2011). *Potential of Using Virtual Environment for Aircraft Maintenance Learning System in Making Tacit Knowledge Explicit*. 506–516. https://doi.org/10.1007/978-3-642-22170-5_43

Jensen, L. X., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, *23*(4), 1515–1529.

Julian Lockwood. (2017). *X-Plane 11: Cessna 172: Pilot's Operating Manual*. Laminar Research.

Kaplan, A. D., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human Factors, 63*(4), 706–726. https://doi.org/10.1177/0018720820904229

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220.

https://www.tandfonline.com/action/showCitFormats?doi=10.1207/s15327108ijap0303_3

Kristin Goodwillie. (2022, November 14). Alaska Airlines one of first US Airlines to use virtual reality in pilot training. *King 5 Media Group*. https://www.king5.com/article/tech/alaska-airlines-use-virtual-reality-pilot-training/281-3cb512f2-6539-41af-8d17-addbdaa9d31d?fbclid=IwAR01Dl1J0NIwu6Klg0TMQZCzOqVvhlDtJVVWx6FMbD0-rgkndnsNt3Ee1po

Kunche, A., Puli, R. K., Guniganti, S., & Puli, D. (2011). Analysis and Evaluation of Training Effectiveness. *Human Resource Management Research*, *1*(1), 1–7. https://doi.org/10.5923/j.hrmr.20110101.01

Lee, K. (2012). Augmented Reality in Education and Training. *TechTrends*, *56*(2), 13–21.

Mansikka, H., Virtanen, K., & Harris, D. (2019). Comparison of NASA-TLX scale, modified Cooper-Harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. *Ergonomics*, *62*(2), 245–254. https://doi.org/10.1080/00140139.2018.1471159

Mazuryk, T., & Gervautz, M. (1996). *Virtual Reality History, Application, Technology and Future*. 1–72.

Michelle P. Hight, Stephanie G. Fussell, Martin A. Kurkchubasche, & Ian J. Hummell. (2022). Effectiveness of Virtual Reality Simulations for Civilian, Ab Initio Pilot Training. *Journal of Aviation/Aerospace Education & Research*, *31*(1), 1–17. https://doi.org/10.15394/jaaer.2022.1903

Mohanty, P. C., Dash, M., Dash, M., & Das, S. (2019). A study on factors influencing training effectiveness. *Revista Espacios*, *40*, 7–15.

Motion System. (n.d.). Redbird Flight Simulations, Inc. Retrieved May 9, 2022, from https://simulators.redbirdflight.com/technologies/motion-system

Myers III, P. L., Starr, A. W., & Mullins, K. (2018). Flight Simulator Fidelity, Training Transfer, and the Role of Instructors in Optimizing Learning. *International Journal of Aviation, Aeronautics, and Aerospace*, *5*(1). https://doi.org/10.15394/ijaaa.2018.1203

Natalia Dużmańska, Paweł Strojny, & Agnieszka Strojny. (2018). Can Simulator Sickness Be Avoided? A Review on Temporal Aspects of Simulator Sickness. *Frontiers in Psychology*, *9*(2132). https://doi.org/10.3389/fpsyg.2018.02132

New US University Flight Training Program Using Virtual Reality Cuts Time To Solo By 30%. (2022, November 25). AFM. https://afm.aero/new-us-university-flight-training-program-using-virtual-reality-cuts-time-to-solo-by-30/?fbclid=IwAR3gqX8WVMXDDeOJZgOCS_7esgDnvspL4ban6OFi9zTLhwbNSC_M61eM2co

Pauline Bimberg, Tim Weissker, & Alexander Kulik. (2020). On the Usage of the Simulator Sickness Questionnaire for Virtual Reality Research. *IEEE Conference on Virtual Reality and 3D User Interfaces, Abstracts, and Workshops (VRW)*, 464–467. https://www.uni-weimar.de/fileadmin/user/fak/medien/professuren/Virtual_Reality/documents/publications/2020-ssq-usage-workshop-wisp.pdf

Prepar3D Product Overview. (n.d.). Prepar3D. Retrieved May 20, 2021, from https://www.prepar3d.com/product-overview/

Qualifications for pilot licences and permits. (n.d.). Government of Canada. Retrieved July 8, 2021, from https://tc.canada.ca/en/aviation/licensing-pilots-personnel/flight-crew-licences-permits-ratings/qualifications-pilot-licences-permits

Quest 2. (n.d.). Oculus. Retrieved May 19, 2021, from https://www.oculus.com/quest-2/

Rongbing Xu. (2022). *Modeling Pilot Flight Performance on Pre-flight and Take-off Tasks with A Cognitive Architecture* [Master of Applied Science in Systems Design Engineering, University of Waterloo]. https://uwspace.uwaterloo.ca/bitstream/handle/10012/18174/Xu_Rongbing.pdf

Rongbing Xu & Shi Cao. (2021). Modeling pilot flight performance in a cognitive architecture: Model demonstration. *Human Factors and Ergonomics Society Annual Meeting*, *65*.

Simulators. (n.d.). Redbird Flight Simulations, Inc. Retrieved May 19, 2021, from https://simulators.redbirdflight.com/products/topic/simulators

Suzanne K. Kearns. (2021). *Fundamentals of International Aviation* (2nd ed.). Routledge.

TCCA and EASA rulemaking co-operation. (2008, January 15). Government of Canada. https://tc.canada.ca/en/aviation/aircraft-airworthiness/international-agreements-arrangements/tcca-easa-rulemaking-co-operation

The Experience. (n.d.). AeroSim Experience. Retrieved May 19, 2021, from https://aerosimexperience.com/lexperience-2/#simulators

Transport Canada. (2019). *Standard 421—Fligth Crew Permits, Licences and Ratings—Canadian Aviation Regulations (CARs)*. Government of Canada. https://tc.canada.ca/en/corporate-services/acts-regulations/list-regulations/canadian-aviation-regulations-sor-96-433/standards/standard-421-flight-crew-permits-licences-ratings-canadian-aviation-regulations-cars#421_26

Transport Canada. (2021). *Flight Test Guide—Private Pilot Licence—Aeroplane—TP 13723 (Sixth Edition—Revised)*. Government of Canada. https://tc.canada.ca/en/aviation/publications/flight-test-guide-private-pilot-licence-aeroplane-tp-13723#g18

Tüzün, İ. K. (2005). General Overview of Training Effectiveness and Measurement Models. *Ticaret ve Turizm Eğitim Fakültesi Dergisi Yıl*, *1*, 144–156.

U. S. Department of Transportation and Federal Aviation Administration. (1995). *Advisory Circular: 120-40C Airplane Simulator Qualification*.

Usability Testing of a VR Flight Training Program. (2021). *Human Factors and Ergonomics Society Annual Meeting*, *65(1)*, 1124–1128. https://doi.org/10.1177/1071181321651096

Velev, D., & Zlateva, P. V. (2017). Virtual Reality Challenges in Education and Training. *International Journal of Learning and Teaching*, *3*(1), 33–37. https://doi.org/10.18178/ijlt.3.1.33-37

Visual System. (n.d.). Redbird Flight Simulations, Inc. Retrieved May 9, 2022, from https://simulators.redbirdflight.com/technologies/visual-system

Vive. (n.d.). Vive. Retrieved May 9, 2022, from https://www.vive.com/ca/

VR Flight Training: Lufthansa Aviation Training & NMY. (2022, February 17). Ultraleap. https://www.ultraleap.com/company/news/case-study/lufthansa-vr-training/

Wallace, J. W., Hu, Z., & Carroll, D. A. (2020, December). Augmented Reality for Immersive and Tactile Flight Simulation. *IEEE Aerospace and Electronic Systems Magazine*, *35*(12), 6–14.

Walter, H., Li, R., Munafo, J., Curry, C., Peterson, N., & Stoffregen, T. (2019). *A brief explanation of the Simulator Sickness Questionnaire (SSQ)*. https://conservancy.umn.edu/bitstream/handle/11299/201892/SSQ%20Scoring.pdf?sequence=9&isAllowed=y

Washburn, C., Stringfellow, P., & Gramopadhye, A. (2007). Using Multimodal Technologies to Enhance Aviation Maintenance Inspection Training. *First International Conference on Digital Human Modeling, ICDHM*, *4561*, 1018–1026. https://link-springer-com.proxy.lib.uwaterloo.ca/chapter/10.1007/978-3-540-73321-8_114

Waterloo Institute for Sustainable Aeronautics. (n.d.). *Research and Partnerships*. University of Waterloo. Retrieved November 23, 2022, from https://uwaterloo.ca/sustainable-aeronautics/research-and-partnerships

William S. Helton. (2004). *Validation of a Short Stress State Questionnaire*. *48*, 1238–1242. https://journals.sagepub.com/doi/pdf/10.1177/154193120404801107?casa_token=9VZEp5WrPYAAAAAA:0Fiy-s71U2OWomxOT_CyonKhGGfeIGgR8OIRrA9sMB58GSliw4TN9i6_jqqQhrM8uPpCczoi9RRk

William S. Helton & Katharina Näswall. (2015). Short Stress State Questionnaire: Factor Structure and State Change Assessment. *European Journal of Psychological Assessment*, *31*(1), 20–30. https://doi.org/10.1027/1015-5759/a000200

Women in Aviation. (n.d.). Brampton Flight Centre. Retrieved October 21, 2022, from https://www.bramptonflightcentre.com/recreation/learn-to-fly/women-in-aviation/

X-Plane 11. (n.d.). X-Plane. Retrieved May 20, 2021, from https://www.x-plane.com/

Yi Gao, Kai Tai Stephen Au, Hyuk Joo Kwon, & Eu Wing Leong. (2013). Learning Styles of Australian Aviation Students: An Assessment of the Impact of Culture. *The Collegiate Aviation Review*, *31*(1), 17–39.

Yuviler-Gavish, N., & Gopher, D. (2013). Task Analysis for Developing Maintenance and Assembly VR Training Simulators. *Ergonomics in Design*, *21*(1), 12–19.

Appendix A
Questionnaires

The Questionnaires used in the study are provided below.

A.1 Screening Questionnaire

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB 44559).

Your participation in the study is voluntary. You may decide to leave the study at any time by communicating this to the researcher. Any information provided up to that point will not be used. You may decline to answer any question(s) you prefer not to answer. You can request that your data be removed from the study up until February 2023 as it is not possible to withdraw your data once my thesis has been submitted.

If you have any questions regarding this study, or would like any additional information regarding your participation, please ask the research team or contact Naomi Paul by email at nvpaul@uwaterloo.ca.

Name: _________

Are you currently a student pilot at the University of Waterloo?

- Yes
- No

Are you currently in the Waterloo/Kitchener Region? If not, can you come to the University of Waterloo main campus to participate in the experiment?

- Yes
- No

Have you completed your first solo flight?

- Yes
- No

How many hours of flight experience do you have with any model of Cessna 172? _________

Please rate how susceptible you are to motion (simulator) sickness?

	1	2	3	4	5	6	7	8	9	10	
Not at all Susceptible	○	○	○	○	○	○	○	○	○	○	Extremely Susceptible

Do you have any of the following medical conditions?

- Seizures
- Medical Devices such as a cardiac pacemaker, hearing aids, and defibrillators
- Diagnosis of a disease of the middle ear or an ear infection in the past 12 months
- Eye, skin, or scalp infection in the past 12 months
- None of the Above
- Other: _________

A.2 Demographic Questionnaire

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB 44559).

Your participation in the study is voluntary. You may decide to leave the study at any time by communicating this to the researcher. Any information provided up to that point will not be used. You may decline to answer any question(s) you prefer not to answer. You can request that your data be removed from the study up until February 2023 as it is not possible to withdraw your data once my thesis has been submitted.

If you have any questions regarding this study, or would like any additional information regarding your participation, please ask the research team or contact Naomi Paul by email at nvpaul@uwaterloo.ca.

Participant ID: _________

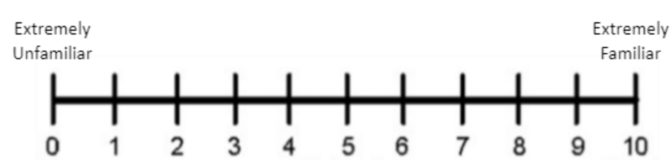
What gender do you identify with?

- Male
- Female
- Two-Spirit
- Non-Binary
- Other: _________

What is your age? _________

What is your best estimate for how many hours of aircraft operation and training you have completed (include simulator training)? _________

Please rate your familiarity with the Region of Waterloo International Airport:



A.3 Simulator Sickness Questionnaire (SSQ)

Pre-Task Questionnaire

Please indicate how well each word describes how you feel *At The Moment*.

Not at all = 1 A little bit = 2 Somewhat = 3 Very much = 4 Extremely = 5					
1. Dissatisfied	1	2	3	4	5
2. Alert	1	2	3	4	5
3. Depressed	1	2	3	4	5
4. Sad	1	2	3	4	5
5. Active	1	2	3	4	5
6. Impatient	1	2	3	4	5
7. Annoyed	1	2	3	4	5
8. Angry	1	2	3	4	5
9. Irritated	1	2	3	4	5
10. Grouchy	1	2	3	4	5

Please indicate how true each statement is of your thoughts *During The Past Ten Minutes*.

Not at all = 1 A little bit = 2 Somewhat = 3 Very much = 4 Extremely = 5					
11. I am committed to attaining my performance goals	1	2	3	4	5
12. I want to succeed on the task	1	2	3	4	5
13. I am motivated to do the task	1	2	3	4	5
14. I'm trying to figure myself out	1	2	3	4	5
15. I'm reflecting about myself.	1	2	3	4	5
16. I'm daydreaming about myself.	1	2	3	4	5
17. I feel confident about my abilities.	1	2	3	4	5
18. I feel self-conscious.	1	2	3	4	5
19. I am worried about what other people think of me.	1	2	3	4	5
20. I feel concerned about the impression I am making.	1	2	3	4	5
21. I expect to perform proficiently on this task.	1	2	3	4	5
22. Generally, I feel in control of things.	1	2	3	4	5
23. I thought about how others have done on this task.	1	2	3	4	5
24. I thought about how I would feel if I were told how I performed.	1	2	3	4	5

Figure 25: SSSQ Pre-Questionnaire (William S. Helton & Katharina Näswall, 2015)

Post-Task Questionnaire

Please indicate how well each word describes how you felt *During The Task.*

Not at all = 1 A little bit = 2 Somewhat = 3 Very much = 4 Extremely = 5					
1. Dissatisfied	1	2	3	4	5
2. Alert	1	2	3	4	5
3. Depressed	1	2	3	4	5
4. Sad	1	2	3	4	5
5. Active	1	2	3	4	5
6. Impatient	1	2	3	4	5
7. Annoyed	1	2	3	4	5
8. Angry	1	2	3	4	5
9. Irritated	1	2	3	4	5
10. Grouchy	1	2	3	4	5

Please indicate how true each statement was of your thoughts *While Performing The Task.*

Not at all = 1 A little bit = 2 Somewhat = 3 Very much = 4 Extremely = 5					
11. I was committed to attaining my performance goals	1	2	3	4	5
12. I wanted to succeed on the task	1	2	3	4	5
13. I was motivated to do the task	1	2	3	4	5
14. I tried to figure myself out.	1	2	3	4	5
15. I reflected about myself.	1	2	3	4	5
16. I daydreamed about myself.	1	2	3	4	5
17. I felt confident about my abilities.	1	2	3	4	5
18. I felt self-conscious.	1	2	3	4	5
19. I was worried about what other people think of me.	1	2	3	4	5
20. I felt concerned about the impression I was making.	1	2	3	4	5
21. I performed proficiently on this task.	1	2	3	4	5
22. Generally, I felt in control of things.	1	2	3	4	5
23. I thought about how others have done on this task.	1	2	3	4	5
24. I thought about how I would feel if I were told how I performed.	1	2	3	4	5

Figure 26: SSSQ Post-Questionnaire (William S. Helton & Katharina Näswall, 2015)

A.4 Subjective Stress Scale

During the following sessions, you will be asked to rate your stress levels during sessions using a 10-point scale. “One” is not stressed at all, similar to a peaceful, relaxing afternoon. “Ten” is the most possible stress you can withstand; for this experiment, that would be the stress you can withstand before asking to terminate the experiment.

Rate your stress on a scale from 1 to 10.

	1	2	3	4	5	6	7	8	9	10	
Not Stressed at All (Peaceful, Relaxing)	○	○	○	○	○	○	○	○	○	○	Most Stress you can Withstand

A.5 NASA Task Load Index (NASA-TLX)

Descriptions of Factors

1. Mental Demand (Low/High): How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
2. Physical Demand (Low/High): How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
3. Temporal Demand (Low/High): How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
4. Performance (Good/Poor): How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
5. Effort (Low/High): How hard did you have to work (mentally and physically) to accomplish your level of performance?
6. Frustration Level (Low/High): How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

Rating Sheet

For each factor, rank your experience while completing a) the before start checklist and b) the takeoff and steep turn:

1. Mental Demand
2. Physical Demand
3. Temporal Demand
4. Performance
5. Effort Frustration

For 1, 2, 3, 5, and 6, rankings use the following scale:

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	○	○	○	○	High

For performance, factor 4, the following scale is used:

	1	2	3	4	5	6	7	8	9	10	
Good	○	○	○	○	○	○	○	○	○	○	Poor

Appendix B
Flight Guides

The flight guides used in the study are provided below.

B.1 Flight Test Guide

The flight test guide used for the instructor evaluations, modified from the official Transport Canada guides (Transport Canada, 2021), is as follows:

Training Device:________________________________ Identifier:______________________________

Pre-flight and Engine Start and After Start					
demonstrate an awareness of other persons and property before and during engine start	1	2	3	4	
use the appropriate checklist provided by the manufacturer or aeroplane owner	1	2	3	4	
accurately complete the engine and aeroplane systems checks	1	2	3	4	
check flight controls for freedom of operation and correct movements	1	2	3	4	
Takeoff					
complete appropriate checklist	1	2	3	4	
check for traffic	1	2	3	4	
advance throttle smoothly to takeoff power	1	2	3	4	
maintain directional control during the takeoff roll	1	2	3	4	
rotate at recommended airspeed of 55 KIAS (+10/-5 knots)	1	2	3	4	
accelerate to and maintain recommended climb speed of 70-80 KIAS (+10/-5 knots)	1	2	3	4	
Steep Turn					
perform and maintain an effective lookout before and during the turn	1	2	3	4	
roll into and out of turns using smooth and coordinated pitch, bank, yaw and power control	1	2	3	4	
roll into a coordinated turn with an angle of bank of 45°	1	2	3	4	
maintain coordinated flight	1	2	3	4	
maintain the selected altitude of 4500 ft (± 100 feet)	1	2	3	4	
maintain airspeed of 100 KIAS (± 10 knots)	1	2	3	4	
maintain 45 degrees angle of bank (± 10°)	1	2	3	4	
visually recover from the turn at the pre-selected recovery reference point (± 10°)	1	2	3	4	

Figure 27: Modified Flight Test Guide for Instructor Evaluations

B.2 Flight Task Checklists

The three flight tasks used for this study include the before start checklist, takeoff, and a steep turn. To complete these procedures properly, the normal operating procedures from the Cessna 172S Skyhawk Information Manual (*172S Skyhawk: Information Manual*, 2004) must be followed.

Due to limitations in the controllers and instrument panels of the flight training device, desktop simulator, and virtual reality simulator, there are minor differences in the procedures followed.

Checklists for Flight Training Device

Before Start

1) Avionics Master Switch – OFF
2) Battery – ON
3) Fuel Level – CHECKED
4) Flaps – UP
5) Throttle – OPEN ¼ INCH
6) Mixture – FULL
7) Strobes – ON
8) Magnetos – ON
9) Fuel Pump – ON
10) Parking Brake – ON
11) Prop Area – CLEAR
12) Starter – ENGAGE
13) Alternator – ON
14) Fuel Pump – OFF
15) Oil Pressure – CHECK
16) Nav Lights – ON
17) Avionics Master Switch – ON
18) Flight Controls – FREE and CORRECT

Takeoff

1) Mixture – FULL RICH
2) Throttle – FULL OPEN
3) Engine Instruments – CHECK
4) Rotate – 55 KIAS
5) Climb Speed – 70-80 KIAS

Checklists for Desktop Simulator

Before Start

1) Ignition Switch – OFF
2) Avionics Master Switch – OFF
3) Master Switches – ON
4) Fuel Level – CHECKED
5) Flaps – UP
6) Throttle – OPEN ¼ INCH
7) Mixture – IDLE CUTOFF
8) Beacon – ON
9) Brakes – SET (Use Toe Brakes)
10) Prop Area – CLEAR
11) Ignition Switch – START (Return to BOTH/ALL when engine starts)
12) Mixture – advance smoothly to FULL RICH as engine starts
13) Oil Pressure – CHECK
14) Nav Lights – ON
15) Avionics Master Switch – ON
16) Flight Controls – FREE and CORRECT

Takeoff

1) Mixture – FULL RICH
2) Throttle – FULL OPEN
3) Engine Instruments – CHECK
4) Rotate – 55 KIAS
5) Climb Speed – 70-80 KIAS

Checklists for Virtual Reality Simulator

Before Start

1) Ignition Switch – OFF
2) Avionics Master Switch – OFF
3) Master Switches – ON
4) Fuel Level – CHECKED
5) Flaps – UP
6) Throttle – OPEN ¼ INCH
7) Mixture – IDLE CUTOFF
8) Beacon – ON
9) Parking Brake – SET
10) Prop Area – CLEAR
11) Ignition Switch – START (Release when engine starts)
12) Mixture – advance smoothly to FULL RICH as engine starts
13) Oil Pressure – CHECK
14) Nav Lights – ON
15) Avionics Master Switch – ON
16) Flight Controls – FREE and CORRECT

Takeoff

6) Mixture – FULL RICH
7) Throttle – FULL OPEN
8) Engine Instruments – CHECK
9) Rotate – 55 KIAS
10) Climb Speed – 70-80 KIAS