7-2024

# Evaluating the Influence of eVTOL Pilot Interface Visual Density and Information Density on Pilot Situation Awareness, Workload, and Search Performance

Bhoomin Bhupendrabhai Chauhan
*Florida Institute of Technology*, bchauhan2017@fit.edu

# Evaluating the Influence of eVTOL Pilot Interface Visual Density and Information Density on Pilot Situation Awareness, Workload, and Search Performance

By

Bhoomin Bhupendrabhai Chauhan

Bachelor of Engineering
Aeronautical Engineering
Gujarat Technological University
May 2015

Master of Science
Applied Aviation Safety
Florida Institute of Technology
May 2019

A dissertation submitted to the College of Aeronautics at
Florida Institute of Technology
as part of the degree requirements for a

Doctor of Philosophy
In
Aviation Sciences

Melbourne, Florida
July 2024

We, the undersigned committee
hereby approve the attached dissertation

Evaluating the Influence of eVTOL Pilot Interface Visual Density and Information Density on Pilot Situation Awareness, Workload, and Search Performance

by

Bhoomin Bhupendrabhai Chauhan


____________________________
Meredith Carroll, Ph.D.
Professor
College of Aeronautics
Major Advisor

____________________________
John E. Deaton, Ph.D.
Professor and Dean
College of Aeronautics

____________________________
Rian Mehta, Ph.D.
Associate Professor and Assistant Dean
College of Aeronautics

____________________________
Brooke E. Wheeler, Ph.D.
Associate Professor
College of Aeronautics

____________________________
Siddhartha Bhattacharyya, Ph.D.
Associate Professor
Electrical Engineering and Computer Science

**ABSTRACT**

TITLE: Evaluating the Influence of eVTOL Pilot Interface Visual Density and Information Density on Pilot Situation Awareness, Workload, and Search Performance

AUTHOR: Bhoomin Bhupendrabhai Chauhan

MAJOR ADVISOR: Dr. Meredith Carroll

The purpose of the current study was to investigate the influence of levels of visual density (VD) and information density (ID) of an electrical vertical take-off and landing (eVTOL) aircraft pilot interface on pilot situation awareness (SA), workload, and search performance. An eVTOL aircraft is a novel aircraft design that is able to perform vertical take-off and landing similar to a rotorcraft and transition to a forward flight, similar to a fixed-wing aircraft. These aircraft are envisioned to operate in urban environments at lower altitudes, necessitating efficient, clear, and concise pilot interfaces to ensure safety and operational effectiveness. The study used a within subjects, quasi experimental design to determine the effect of varying levels of VD and ID on the pilot's SA, workload, and search performance. The sample consisted of 26 instrument-rated student pilots, who performed four trials involving landing an eVTOL aircraft at four airports using varying levels of VD and ID.

A multivariate analysis of variance (MANOVA) revealed that there was a significant effect of levels of VD and ID on SA, workload, and search performance.

Further analysis also revealed a significant interaction between VD and ID on search performance. The findings of the study were consistent with the SEEV model (Wickens et al., 2001) and Broadbent's (1958) filter model of attention. The study's findings align with research in the aviation domain and provide strong evidence that the levels of VD and ID can affect an eVTOL pilot's SA, workload, and search performance.

## Table of Contents

## List of Figures

# List of Tables

**Chapter 5**

**Acknowledgment**

As an old proverb goes, “If you want to go fast, go alone. If you want to go far, go together.” This dissertation would not have been possible without the care and support of several people, who have directly or indirectly made a lasting impact on this work and me.

First of all, thank you, Dr. Carroll. I do not know of any form of gratitude that I can summon that could sufficiently express how grateful I am to get the opportunity to work under your tutelage. This dissertation would not have been possible without your tireless effort, patience, and guidance. I am forever grateful for giving me the opportunity to pursue this work, particularly in advanced air mobility. Since the first conference paper, the experience I have garnered, and the network I have built, it would not have been possible without your help. Thank you to my committee members: Dr. Mehta, Dr. Wheeler, and Dr. Bhattacharyya. Thank you for your valuable input and for ensuring the quality of this work is the best it can be. Your input has not only improved the quality of my work but also helped me grow as a researcher.

Thank you to my friends at the ATLAS Lab. Thank you for helping me whenever I was stuck on a problem. Thank you for all the brainstorming sessions, discussions, and tips. Lastly, this dissertation would not have been possible without two people whom I am lucky to call my closest friends: Avani and Rosmy. Thank you for listening to me for hours on end, motivating me, believing in me, and pushing me to give my best.

## Dedication

I would like to dedicate this dissertation to my mother, father, and sister.

# Chapter 1

# Introduction

## Background and Purpose

### *Purpose*

The purpose of the current study was to examine the effect of varying levels of visual density (VD) and information density (ID) of a simulated Beta ALIA-250, an electric vertical take-off and landing (eVTOL) aircraft pilot interface on situation awareness (SA), workload (NASA-TLX; Hart, 1986), and search performance during a simulated landing phase of the flight on a desktop-based flight simulator. VD is defined as a surplus of items, information, or the number of objects within a display (Horrey & Wickens, 2004) and was manipulated by adding and removing customizable pieces of information from the testbed pilot interface. ID, in the current study, is defined as the ratio of the total quantity of relevant information to the total quantity of information on the testbed pilot interface (Alexander et al., 2009). ID was manipulated by adding and removing customizable relevant, irrelevant, and redundant information on the testbed pilot interface. An eVTOL aircraft is a novel aircraft concept capable of vertical take-off and landing using an electrical propulsion system, in this study, the option available in X plane 12 was the BETA Technologies ALIA 250 aircraft. Participants were tasked to complete a total of four approaches using the simulated eVTOL aircraft at four major airports using a flight simulation software with four display conditions with varying levels of VD and ID. The dependent variables were SA (measured using SAGAT queries; Endsley, 1995), workload (measured using NASA-TLX; Hart, 1986), and search performance (measured by the number of seconds taken to identify the final approach fix waypoint). Definitions for the target variables, relevant key terms, and variables are provided later in Chapter 1.

***Background***

As a result of new technological advances in the eVTOL aircraft design, the advent of electric propulsion systems, newer and more efficient battery technology, and advanced aircraft automation, an increasing number of aviation stakeholders have been investigating the integration of advanced air mobility (AAM; Cohen et al., 2021). AAM is a novel form of air transportation that envisions moving passengers and freight in urban, suburban, and rural regions (Andritsos et al., 2022; Cohen et al., 2021). From a passenger perspective, AAM includes commutes within a 50-mile radius and intraregional use cases for up to 100 miles using an eVTOL aircraft (Goyal et al., 2021). AAM envisions integrating crewed eVTOL flights (in the near term) and fully autonomous flights (in the long term) of varying mission classifications (for example, passenger, cargo, and medivac) in the National Airspace System (NAS; National Academies of Sciences, Engineering, and Medicine, 2020). According to the latest Federal Aviation Administration (FAA) concept of operations (ConOps), initial AAM flights will have a pilot on board, and AAM operators, who will offer eVTOL flights, will be required to operate under the 14 Code of Federal Regulations (CFR) Part 135 (FAA, 2023b, 2023d). For these early adopters of eVTOL operations, the FAA expects that pilots will be required to operate from the surface to 4000’ above ground level (AGL) and adhere to the established two-way radio communication, navigation, and surveillance (CNS) requirements of the airspace they are operating in (FAA, 2023b). In addition, due to the unavailability of dedicated ground infrastructure, also known as vertiports, in urban and metropolitan areas, eVTOLs will predominately operate in relatively close proximity to, or directly on, airports in or around Class B and Class C Airspace (Mendonca et al., 2022).

According to a report published by the Government Accountability Office (GAO), one of the critical components of the AAM ecosystem is the eVTOL aircraft and the pilots that will fly these aircraft (GAO, 2022; Lineberger et al., 2019). Currently, eVTOL aircraft are not certified for commercial operations, and there is no established pilot certification requirement. As these novel aircraft continue to develop, the FAA has proposed alternate eligibility criteria to expedite the pilot certification process. This accelerated process would apply to pilots with a commercial pilot license (CPL) and an instrument rating (FAA, 2023c). While the operational ecosystem for near–term AAM flights will be similar to fixed-wing and rotorcraft operations, one of the primary differentiators between a conventional rotorcraft/aircraft and eVTOL aircraft will be the autonomous capability of the eVTOL aircraft, use of electric propulsion, and the pilot interface that the pilots will use to fly these aircraft. The eVTOL pilot interface will be different from the conventional aircraft cockpit in that the pilot interface would be required to present the advanced autonomous capability of the aircraft, and the real estate available in the cockpit to display information will be much more restricted compared to conventional aircraft and rotorcraft (Bacchini & Cestino, 2019). Due to the limited space available in the cockpit, eVTOL original equipment manufacturers (OEMs) have partnered with companies like Garmin, Avidyne, and Honeywell to develop customized pilot interfaces for their eVTOL aircraft (Archer Aviation, 2022; Alcock, 2021; Garmin, 2021; Wyrick, 2023). These eVTOL pilot interfaces will be tailored to the eVTOL aircraft and display flight–critical information in the limited space in the cockpit, with customization to allow the pilot to decide which information they need during different phases of the flight.

Modern aircraft and helicopter cockpits have evolved from multiple analog displays to glass and touch cockpit displays. A typical commercial, passenger-service aircraft flight deck has up to six multi-function displays (MFDs), backup flight instruments, and several critical system indicators on the main instrument panel (Zhang et al., 2014). Generally, aircraft system controls are located on an overhead systems panel, and a mode control panel, also referred to as a flight control unit, is located centrally on the glare shield below the windscreens. Other flight management system (FMS) controls, communication controls, aircraft power system controls, and configuration controls are located on the pedestal between the pilots (Zhang et al., 2014). However, the exact positioning of these displays and what information is presented can vary significantly between different aircraft types and the air carrier’s requirements. This is why pilots are required to get a separate type rating for each aircraft model they operate. The eVTOL pilot interfaces currently under consideration for use in AAM-supporting aircraft have several key differences from the traditional pilot interfaces. For example, an eVTOL pilot interface will likely consist of a single or dual MFD setup, integrate electric propulsion system information (e.g., multiple rotor rotation per minute information, battery level), powerplant setup, and unique flight characteristics, for example, automated vertical take-off (AOPA, 2022; Courtin et al., 2021; Kinjo, 2018).

One clear distinction between a traditional aircraft pilot interface and an eVTOL pilot interface is that there will be fewer displays. Given the amount of space available and other drivers, such as the desire to reduce pilot training requirements, current efforts are focused on simplifying the displays by reducing the redundant and non–critical flight information, resulting in a pilot interface that is comparatively less cluttered than a traditional pilot interface. While clutter on a visual display and its different measures (e.g., display layout, display density, and task-relevant information) have been extensively studied in aviation, both for commercial and military applications, research from an eVTOL pilot interface perspective is limited and warrants further examination. Several bodies of research have studied the impact of pilot interface clutter and have shown that both high and low levels of clutter can degrade pilot situation awareness (SA; Endsley, 1988; Moacdieh et al., 2013; Wu et al., 2016), increase pilot workload (Doyon-Poulin et al., 2014; Hoh et al., 1987), and result in lower flight performance (Doyon-Poulin et al., 2012; Moacdieh et al., 2013). However, human factors research for an eVTOL aircraft is sparse. Much of the current eVTOL research is focused on improving the eVTOL aircraft design, commercialization prospects, and automation (Vempati et al., 2023). Limited research is currently available to the research community concerning the influence of the eVTOL pilot interface on pilot SA, workload, and search performance. Considering the rate at which the AAM industry is developing, OEMs actively developing eVTOL aircraft are, no doubt, conducting internal human factors research; however, the competitive nature of the industry has led to minimal OEM research findings being made available to the public, academia, or interested stakeholders due to the fear of sharing technical and intellectual details regarding their eVTOL aircraft. Further, OEMs are trying to simplify the eVTOL pilot interface by removing redundant and non-critical information. Therefore, academic research examining the effect of various visual display clutter measures, such as VD and ID of a pilot interface and the pilot's ability to use the information on an eVTOL pilot interface to conduct a mission safely, is needed.

To ensure the safety of both passengers and cargo, eVTOL pilots must be able to use the information displayed on the pilot interface effectively. Due to the uncertainty around information that will be presented on an eVTOL pilot interface, including the information display characteristics, the quantity, and the presence of relevant information, it is crucial to examine how various characteristics of an eVTOL pilot interface, particularly the level of clutter, will influence the pilot's SA, workload, and search performance. While several characteristics of a visual display have been employed to quantify the level of clutter, VD and ID are the most commonly adopted measures of clutter (Moacdieh & Sarter, 2015).

The VD of a visual display is often described as the total quantity of information within a display (Horrey & Wickens, 2004). In the context of the current study, a display with high levels of VD would have a higher quantity of information. Past research, both in the context of visual displays, in general, and pilot interfaces, has shown that high levels of VD can lead to lower performance (Backs & Walrath, 1992; Bennett et al., 2021; Wickens et al., 2005) and decreased SA (Alexander & Wickens, 2005; Wickens et al., 2004).

ID, another widely used measure of clutter, is described as a high percentage of task-relevant information within a display (Doyon–Poulin et al., 2012). In the current study, a display is considered to have high levels of ID when there is a higher quantity of task-relevant information compared to irrelevant and redundant information, resulting in a higher ratio. Conversely, a display with a low ID would have a higher quantity of redundant and irrelevant information than relevant information, resulting in a low ratio compared to a high ID. Research has shown that a pilot experiences lower workload, higher SA, and improved performance with high ID compared to low ID (Alexander et al., 2003; Morphew & Wickens, 1998; Prinzel et al., 2018).

The current study intended to examine the influence of VD and ID as measures of eVTOL pilot interface clutter on the pilot's SA, workload, and search performance. While previous research has been focused on evaluating the influence of VD and ID on a fixed-wing and rotorcraft pilot interface, there is no published academic research currently available to inform the industry about the influence of VD and ID in eVTOL pilot interfaces. The primary reason for the lack of understanding of this emerging research may be attributed to the lack of access to new eVTOL displays and aircraft, given the proprietary nature of the aircraft and the fact that major OEMs are in their flight test campaign and applying to certify their aircraft with the FAA. Although individual manufacturers may be evaluating the effectiveness of their displays independently, these results are not available to the public, or academia, and lack of access to these interfaces prevents the unbiased academic research that can help inform the broader AAM community. The current study aimed to inform the aviation community by examining the impact of reduced information quantity and redundancy on pilot SA, workload, and search performance.

**Definitions of Terms**

This section presents the definitions of key terms that will be used in the current study.

1. *Advanced air mobility (AAM)* refers to a rapidly emerging sector within the aerospace industry, focused on safely and efficiently integrating novel aircraft into the NAS (National Academies of Sciences, Engineering, and Medicine, 2020). In the current study, to simulate AAM operations, the participants performed four approaches at four

major airports, using existing air traffic procedures on a desktop-based flight simulator.

2. *Critical information* refers to any piece of information that is essential for the pilot to complete a task at hand (Jonsson & Ricks, 1995). In the context of the current study, critical information for an eVTOL aircraft was determined based on the pieces of information deemed critical by SMEs in aviation, with experience working in AAM space.
3. *Clutter* refers to a display that has an excess quantity of unwanted or unnecessary information or is presenting an abundance of irrelevant or redundant information for the task at hand (Ahlstrom, 2005; Doyon-Poulin et al., 2012; Lohrenz et al., 2009). In the current study, the level of clutter was established by varying the levels of VD and ID on the pilot interface of the simulation testbed.
4. Electric vertical take-off and landing *(eVTOL) aircraft* refers to a type of aircraft that utilizes electric propulsion systems to power multiple vertical lift propellers or rotors, enabling it to take off and perform a near–vertical landing (Pavel, 2022). In the current study, participants used a simulated eVTOL aircraft selected for the X plane 12 testbed to complete the task using a desktop-based flight simulator.
5. *Information density (ID)* is defined as the ratio of the total quantity of relevant information to the total quantity of information on a primary flight display (PFD) and MFD (Alexander et al., 2009). In the current

study, ID was calculated by deriving the ratio of relevant information to the sum of redundant, irrelevant, and relevant information, i.e., the total quantity of information. Low and high ID was manipulated by adding and removing customizable relevant, irrelevant, and redundant pieces of information from the simulation testbed pilot interface.

6. *Irrelevant information* refers to any piece of information that does not assist the pilot in completing the task at hand (Doyon-Poulin et al., 2012). For the current study, the quantity of irrelevant information was determined based on the individual pieces of information that were deemed irrelevant for an eVTOL pilot performing an approach at an airport by pilot SMEs, some of whom had experience working in AAM.
7. *Non-critical information* refers to any piece of information that is helpful but not required for the pilot to complete the task at hand. In the current study, non-critical pieces of information were identified by consulting with pilot SMEs, some of whom had experience working in AAM.
8. *Part 135 operator*, refers to an air carrier operator that will be FAA-certified to offer eVTOL flights operating under Part 135 (FAA, 2020). In the current study, to simulate Part 135 operations, participants were tasked to land the simulated eVTOL aircraft at four airports using a flight simulator testbed.
9. *Search Performance*, in the context of the current study, refers to the time taken by a participant to locate a piece of information from the pilot

interface of their aircraft. In the current study, search performance was quantified by measuring the time in seconds for the participants to name the final approach fix waypoint to the runway of the airport, where they were performing the near vertical landing.

10. *Pilot interface*, refers to the PFD and the MFD panels of the flight simulator testbed. The pilot interface was used to present the flight information and develop the display conditions for the experimental manipulations.
11. *Powered-Lift* refers to an aircraft that can change the direction of the thrust generated from the aircraft's propulsion system, both on land and while in flight (FAA, 2023c). In the context of the current study, this operational characteristic makes an eVTOL aircraft different from a traditional, fixed-wing aircraft, and therefore, a simulated eVTOL aircraft was utilized for the study.
12. *Redundant information* refers to any information presented multiple times on the pilot interface (FAA, 2014). In the context of the current study's manipulation, any pieces of information that were presented multiple times on the pilot interface were labeled as redundant pieces of information.
13. *Relevant information* refers to any information that is useful or needed for the pilot to complete a particular task at hand. In the context of the current study, the quantity of relevant information for an eVTOL pilot

interface was determined based on pieces of information that were deemed relevant by pilot SMEs, some of whom had experience working in AAM and with eVTOL aircraft.

14. Situation awareness (*SA)* refers to an individual's perception and understanding of the elements and dynamics within their environment or a specific situation, along with their comprehension of the implications and potential future developments (Endsley, 1995). Participant SA was measured by calculating the sum of correct responses to Level 1, Level 2, and Level 3 Situation Awareness Global Assessment Technique (SAGAT) queries (Endsley, 1998a).
15. *Visual density* refers to a measure of clutter as measured by the total quantity of information presented on the pilot interface (Alexander et al., 2018). In the current study, the VD of the testbed pilot interface was calculated by summing the total number of pieces of information on the pilot interface. VD was manipulated by adding and removing customizable pieces of information from the testbed pilot interface. A low VD display condition had lower quantity of information compared to high VD display condition.
16. *Workload* refers to the cumulative cognitive and physical effort required by the pilot to meet the demands of a specified flight task (Roscoe & Ellis, 1990). The current study measured the participant's workload using the NASA TLX questionnaire (Hart, 1986; see Appendix B).

## Research Questions and Hypotheses

### *Research Questions*

As the intention of the study was to examine the impact of varying levels of VD and ID on the pilot's SA, workload, and search performance, the corresponding experimental research questions that guided the study are as follows:

**Research Question 1.** What is the effect of pilot-interface VD on pilot SA, workload, and search performance?

**Research Question 2.** What is the effect of pilot-interface ID on pilot SA, workload, and search performance?

**Research Question 3.** What is the interaction between the levels of VD and ID with respect to the pilot's SA, workload, and search performance?

### *Research Hypothesis*

The corresponding research hypotheses are as follows:

**Hypothesis 1a**: Pilot interfaces with high VD will lead to lower SA than pilot interfaces with low VD.

**Hypothesis 1b**: Pilot interfaces with high VD will lead to a higher workload than pilot interfaces with low VD.

**Hypothesis 1c**: Pilot interfaces with high VD will lead to lower search performance than pilot interfaces with low VD.

**Hypothesis 2a:** Pilot interfaces with high ID will lead to higher SA than pilot interfaces with low ID.

**Hypothesis 2b:** Pilot interfaces with high ID will lead to a lower workload than pilot interfaces with low ID.

**Hypothesis 2c:** Pilot interfaces with high ID will lead to higher search performance than pilot interfaces with low ID.

**Hypothesis 3a.** There will be an interaction between VD and ID on SA such that when VD is low, high levels of ID will result in increased SA, but when VD is high, higher levels of ID will cause a decrease in levels of SA.

**Hypothesis 3b.** There will be an interaction between VD and ID on workload such that when VD is low, high levels of ID will result in a lower workload, but when VD is high, higher levels of ID will cause a higher workload.

**Hypothesis 3c.** There will be an interaction between VD and ID on performance such that when VD is low, high levels of ID will cause an increase in search performance, but when VD is high, higher levels of ID will lead to lower search performance**.**

### Exploratory Qualitative Research Question

To understand the impact of varying levels of VD and ID of an eVTOL pilot interface, a phenomenological research question was also employed to guide the study:

#### *Research Question 4*

What is the participant's reaction to using each of the four display conditions?

Participant reactions were collected using a series of open-ended qualitative questions. These responses were not analyzed as part of the primary analysis but were analyzed from an exploratory perspective and presented to the reader.

**Study Design**

The current study utilized a repeated measures research design. This method was ideal as it allowed me to collect quantitative data regarding the impact of varying levels of VD and ID on pilot SA, workload, and search performance and collect subjective reactions regarding the usability and reaction to using varying levels of eVTOL pilot interface VD and ID. For the experimental component, a within-subjects repeated measures design was used with two independent variables: VD (low vs. high VD ) and ID (low vs. high ID). This research methodology allowed me to identify statistical disparities between the different levels of VD and ID and any potential interactions between the two variables. The study involved participants with diverse backgrounds and varying levels of experience. A within-subject design was utilized to account for individual differences, with all participants experiencing each treatment condition. Consequently, there was a single group of participants in the study.

For the qualitative component, a phenomenological approach was utilized by having the participants respond to a series of open-ended questions. The questions focused on gathering subjective responses about the usability of the eVTOL pilot interfaces in each display condition.

**Significance of the Study**

Several industry stakeholders, regulatory authorities, and OEMs are actively working toward developing and identifying certification requirements for eVTOL aircraft, ConOps, and operational requirements of eVTOL aircraft. Although there is a plethora of aviation display literature that current OEMs can consult, there is still limited academic research regarding how various display characteristics of an eVTOL pilot interface can impact performance and safety. This study is one of the first to contribute to this body of research by examining how eVTOL pilot interface VD and ID impact pilot SA, workload, and search performance. The findings of the current study can help OEMs understand the implications of VD and ID impact on pilot SA, workload, and search performance. Consequently, it can help OEMs develop more effective interfaces, resulting in safer and more efficient eVTOL operations. Considering that the current pilot interfaces enable the pilots to customize which information is presented, the findings of the current study might also help inform guidance for this customization or requirements that will constrain the customizability of the displays. The findings of this study can provide evidence-based recommendations to assist the FAA in certification requirements, ensuring that eVTOL pilot interface designs align with human factors principles.

## Study Limitations and Delimitations

### *Limitations*

Limitations are the conditions, events, and circumstances that are beyond the researcher's control and, therefore, can affect the generalizability of the study and the findings. Limitations associated with the current study include:

**1. Representativeness of the Sample.** The sample consisted of Florida Tech flight students, who hold an instrument rating. Given that the requirements for future eVTOL pilots do not currently exist, and only provisional pilot training requirements have been made available by the FAA, there may be different training requirements in the near future, yielding additional differences between the proposed sample and the eVTOL pilots, limiting the generalizability of the study.

**2. Representativeness of the Scenarios.** In the current study, the experimental tasks that the participants performed were based on the review of the FAA AAM Implementation Plan, FAA's Urban Air Mobility (UAM) ConOps (FAA, 2023d) and recommendations from subject matter experts (SMEs) in aviation with expertise in AAM, aviation planning, air traffic control, and airport operations. As eVTOL aircraft are not certified for commercial operations, the industry does not expect to see commercial AAM flights for at least the next three to four years. Therefore, modifications in factors relative to the flight, departure, and destination sites can change after the current study is concluded. The experimental task, the flight path, and eVTOL using an active landing runway may not represent future eVTOL flights. This limits the generalizability of the current study. Therefore, future studies that utilize scenarios, such as established AAM flight corridors and vertiports, may yield different results from the current study.

**3. Experience in flying an eVTOL aircraft**. In the current study, participants were tasked to fly a simulated eVTOL aircraft in a flight simulator. As eVTOL aircraft are not yet certified by the FAA for commercial operations or flight training, the sample population will not have any experience flying an eVTOL aircraft, which limits the extent of tasks that I can ask the sample population to perform. Any future study that utilizes certified eVTOL pilots or student pilots training to become eVTOL pilots could yield different results.

**4. Relevant versus irrelevant information**. In the current study, I was limited in identifying relevant and irrelevant information for each of the conditions based on the information that was already displayed on the simulator pilot interface and/or the information that could be customized. Additionally, SMEs were consulted to help determine the relevancy and irrelevancy of the information. A different study that utilizes a different pilot interface or uses different SMEs or a different process to establish relevant and irrelevant information, might yield different results.

### *Delimitations*

A study's delimitations are conditions or events that a researcher imposes to make the study more feasible to implement. However, the reader should keep in consideration that these delimitations may further reduce the generalizability of the results. Potential delimitations of the current study include:

**1. Sample Strategy.** The current study utilized convenience sampling with the criterion of completion of instrument rating. Using this screening criterion should allow for control of learning effects and form a homogenous group. As it is still unclear the number of flight hours an eVTOL pilot would require to be certified to fly an eVTOL, a study in the future that uses participants, who are in eVTOL flight training or are eVTOL pilots, may yield different results.

**2. eVTOL Pilot Interface.** The current study utilized a Garmin G1000 pilot interface that was available with the simulation testbed setup. As stated previously, several eVTOL OEMs have proposed using different Garmin display models, for example, Garmin G3000, for their respective eVTOL aircraft. However, the interface chosen for this study may not accurately represent the pilot interface from a certified eVTOL aircraft but presents information that will be included in an eVTOL aircraft. A study in the future that employs a different pilot interface, for example, an Avidyne, a Honeywell pilot interface, or a Garmin G3000 pilot interface, may yield different results.

**3. Simulated eVTOL Aircraft.** The simulated eVTOL was selected for the current study as it was one of the only available fully functional eVTOL aircraft offered by an off-the-shelf flight simulator. The simulated eVTOL aircraft is an accurate model of an actual eVTOL currently being developed for different applications in the AAM ecosystem. While there are other flight simulators, the X-Plane 12 testbed was selected due to the availability of an eVTOL aircraft. However, other simulation testbeds with simulated eVTOL aircraft are available for purchase, none of which are representative of an actual eVTOL aircraft. A study employing a different eVTOL aircraft or testbed may yield different results.

**4. Representativeness of the Scenario Challenges.** The experimental tasks developed for the current study scenarios do not span the full range of flight profiles an eVTOL aircraft would be flying. The scenarios for the current study were developed considering the FAA recommendations for initial AAM operations. Studies that use different sets of scenarios that accurately represent the AAM ecosystem, for example, landing at a vertiport in a metropolitan area, might yield different results.

**5. Representativeness of performance measure.** In the current study, I measured each participant's search performance by measuring the time it took them in seconds to name the final approach fix waypoint using the pilot interface. This search performance measure was selected based on past conventional fixed-wing aircraft literature and results from the data analysis pilot run. However, a different study in the future that utilizes a different performance metric or uses a pilot interface from a certified eVTOL aircraft as a measure of search performance could yield different results.

**6. Independent Variable Manipulation.** The current study manipulated the IVs by adding and removing select customizable pieces of information from the pilot interface to develop the four display conditions. Using the available customizability of the panels allowed for a realistic representation of an eVTOL pilot interface. However, a study that utilizes a different method to manipulate VD and ID may yield different results.

**7. SAGAT Queries.** The current study employed queries designed specifically for this mission and simulator context. These queries were developed using the method outlined by Endsley (2000), combined with previously published task analyses and queries, resulting in a total of 20 queries. However, the limited number of queries and the brief task duration may prevent a comprehensive assessment of SA. Additionally, since the queries were self-developed for this study, they have not undergone extensive testing to ensure their validity and reliability. Consequently, the queries may not have provided the most accurate and robust measure of objective SA, and future studies using a different set of queries may yield different results.

**8. Workload Measure.** The current study used the NASA-TLX questionnaire to measure workload. A different study that uses different measures of workload, such as The Bedford Workload Scale (Roscoe, 1984), or physiological measures, such as cardiovascular activity: Heart Rate (HR), Heart Rate Variability (HRV), and Electrocardiography (ECG) may yield different results.

# Chapter 2

# Literature Review

## Introduction

The current chapter is divided into three sections. The first section will provide a detailed overview of the theories in which the current study was grounded, i.e., the Salience, Effort, Expectancy, and Value (SEEV) Model by Wickens et al. (2001) and Broadbent's Filter Model (1958). The first section will provide a detailed description of the SEEV Model and Broadbent's Filter Model and their relevance to how top-down and bottom-up contributors of the clutter of eVTOL pilot interfaces can influence pilot, SA, workload, and search performance. The second section will provide a detailed review of past research conducted to understand the influence of clutter and its associated top-down and bottom-up contributors, i.e., VD and ID, on SA, workload, and search performance. The third section will summarize the past research findings and their implications for the current study.

## Overview of the Underlying Theories

Clutter has been defined as "unwanted or unnecessary information" (Lohrenz et al., 2009, p. 90), "redundant information" (Ahlstrom, 2005, p. 90), or "an abundance of irrelevant information" (Doyon-Poulin et al., 2012, p. 2D1–2). Research has shown that clutter can degrade monitoring (Schons & Wickens, 1993), delay visual search (Henderson et al., 2009; Neider & Zelinsky, 2011), and negatively affect SA (Kim & Kaber, 2009). These results highlight that clutter can impose significant challenges for an operator to perceive information from their visual stimuli.

Alexander et al. (2008), based on the characteristics of a display, emphasized various aspects of clutter in their research by making the distinction between the bottom-up or data-driven property of the display (i.e., visual/display density, physical appearance, eccentricity; proximity) and the top-down or knowledge-driven property of the display (i.e., ID, relevancy, and redundancy) aspects of clutter. This distinction clearly highlights how both the inherent characteristics of how the information is presented and how the operators process the information are influenced by levels of clutter.

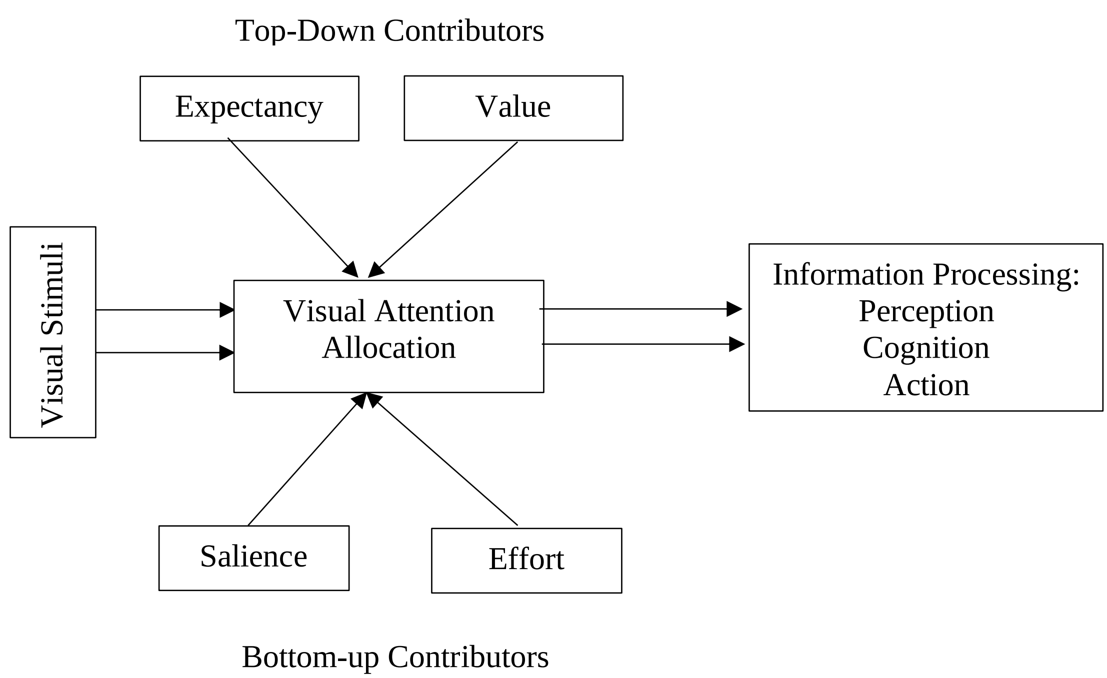
The total quantity of items on display is one of the most frequently adopted and accepted measures of clutter (Horrey & Wickens, 2004). The number of objects presented on a display varies with the domain. For example, in aviation, VD is a measure of the number and proximity of icons, symbols, or pointers on a cockpit display (Kim & Kaber, 2009; Lohrenz et al., 2009; Wickens & Andre, 1990; Yeh & Wickens, 2001). In the current study, the influence of varying levels of VD was characterized as a bottom-up contributor to clutter. Varying levels of ID were characterized as a top-down contributor to clutter and were grounded using the SEEV Model and Broadbent's Filter Model.

***SEEV Model***

Wickens et al. (2001) proposed and evaluated a model of attention that states that the ability to allocate attention is not just limited to the prediction of eye movement as historically proposed by visual attention theories (Carbonell, 1966; Moray, 1986; Senders, 1964) but also by the characteristics of the information that is presented. Wickens et al. (2001) identified four key features of visual stimuli that can influence attention in a dynamic, visually stimulating environment, for example, an aircraft cockpit. As shown in Figure 2.1, these four factors comprise the top-down and bottom-up processes that can influence how attention is directed towards stimuli presented to the operator. These four factors are 1) **S**aliency of the information, 2) **E**ffort needed to undertake (i.e., the time spent locating information by the terms of longer or smaller scans) to identify information, 3) **E**xpected location of the information, and 4) **V**alue of the information (relevant or irrelevant information).

**Figure 2.1**

*SEEV Model*



*Note.* Components of SEEV Model. From “NT-SEEV: A Model of Attention Capture and Noticing on the Flight Deck by C. Wickens, J. McCarley, and K. Steelman-Allen, 2009. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *53*(12), 769–773. (https://doi.org/10.1177/154193120905301202). 

The SEEV Model describes the nature of the underlying attention mechanism that drives visual search in a dynamic environment. According to the model, visual attention is allocated based on bottom-up and top-down information processing when a visual stimulus is presented to the operator. In the bottom-up process, according to the SEEV Model, salience and effort are identified characteristics of a piece of information that help predict how information is captured. Salience can be regarded as the extent to which a piece of information captures attention based on its physical qualities (Schriver & Rantanen, 2007). For example, on a pilot interface, a more salient piece of information will be attended to first compared to other pieces of information that are not as saliently presented. In the model, effort is defined as the extent to which the pilot has to shift their attention to locate information on the pilot interface (Wickens et al., 2001). In the context of the current study, if the information is presented in a display with a higher quantity of information, it will result in a visually dense display, even if it presents task-relevant information. In a visually dense display, the pilot will be required to exercise additional effort within the target pilot interface to access information.

Schriver and Rantanen (2007) stated that the more effort required to access information, the more likely the pilot will spend additional cognitive resources to access the information.

In the top-down process, according to the SEEV model, both expectancy and value parameters of the visual stimuli can help predict how the information will be captured, where the information is expected to be displayed, and determine the value (relevancy) of the information. In the SEEV model, expectancy is defined as the attention allocated to sources of higher task-relevant information. The expectancy in the model describes how the information (change) will appear on the display and how it can influence whether the pilot can focus the attention on the changing information. Through the model, Wickens et al. (2001) identified that an operator is drawn to areas on a display where there is a tendency for frequent change in information. For example, a pilot's attention will be drawn towards the changing altitude during take-off rather than radio frequencies. Value in the SEEV model is defined as the attention allocated to sources of information more valuable (relevant information) to the task. The value of information in a dynamic environment is very task-specific. Any information presented on the display that is irrelevant or redundant will not be captured by the pilot and will be filtered out.

In contrast, a visual search will only capture the task-relevant information. Both expectancy and value highlight that not every piece of information displayed on a pilot interface will aid the pilot in performing the task at hand. Research has shown that low ID, i.e., a higher presence of irrelevant and redundant information, can increase search time (Henderson et al., 2009; Neider & Zelinsky, 2011; Ververs & Wickens, 1998), which means that the pilot will have to spend more time identifying task-relevant information from the redundant and irrelevant information. From all the information displayed on the pilot interface, the pilot has to filter out irrelevant and redundant information. Broadbent's Filter Model explains this filtering process of irrelevant and redundant information.

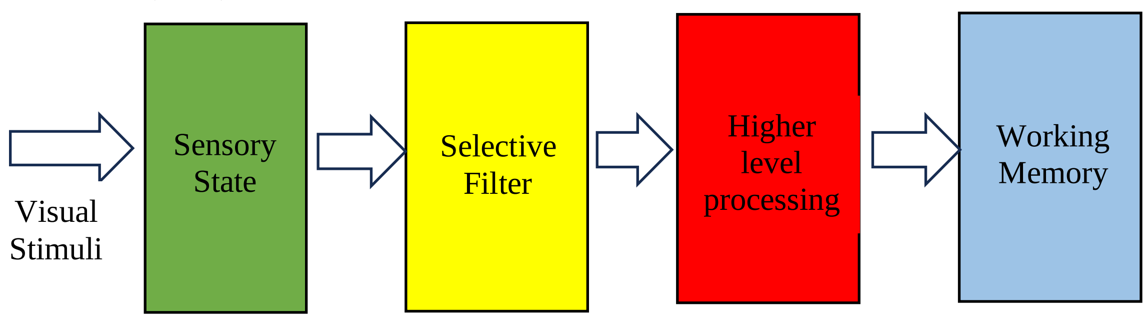
***Broadbent's Filter Model of Attention***

Based on the definition of clutter, ID is not only influenced by the quantity of the information on the display, but also associated with the presence of relevant information available for the task (Doyon-Poulin et al., 2014). Woodruff et al. (1988) initially aimed to assess ID by altering specific information characteristics. However, subsequent studies, such as those by Alexander et al. (2008) and Rosenholtz et al. (2007), adopted different strategies like presenting superfluous details or unrelated task information to manipulate ID. In the current study, ID as a contributor to clutter was theoretically grounded using Broadbent's Filter Model of Attention.

In a dynamic environment, humans are exposed to a variety of stimuli, some of which are relevant to the task at hand, and some that are deemed irrelevant. Broadbent (1958), in his Filter Model, suggested that all stimuli are processed simultaneously based on fundamental physical properties (color, orientation, saliency). In the process of filtering information, Broadbent (1958) argued that when a stimulus is presented, it is first stored in the sensory store, after which the information is passed to a filter, which is regarded as the selector of relevant information. Information relevant to a particular task will be attended to, while irrelevant information is discarded. Research has shown that if irrelevant and redundant information is presented to the pilot, it can severely limit the pilot's visual search performance (Doyon-Poulin, 2014; Moacdieh & Sarter, 2015). The pilot will use up their cognitive resources trying to process and filter out irrelevant and redundant information, resulting in lower SA, increased workload, and lower search performance.

**Figure 2.2**

*Broadbent's (1958) Filter Model*




*Note.* Information filtering process. From Broadbent, E.D. (1958) *Perception and Communication*. Pergamon Press.

Broadbent (1958) proposed a stage model of perception, as shown in Figure 2.2. Based on this model, the initial processing involves analyzing all incoming stimuli to extract fundamental physical characteristics. After the information is made available, Broadbent (1958) stated that information is stored in the sensory store based on the physical characteristics of the stimulus. After the information is stored in the sensory state, a selective filter is used to differentiate certain stimuli to pass through the filter for further processing. In the selective filter model, irrelevant and redundant stimuli are filtered out and lost, while relevant information moves to the next stage for high-level processing, where it is stored in the working memory to be used in the context of the task at hand.

In the current study, VD represented the total quantity of information presented on the pilot interface. The VD of the information on the pilot interface will influence how well the pilot can extract and then perceive information for the task at hand. Based on the SEEV Model, high VD (i.e., the total quantity of information presented on the pilot interface) on the pilot interface will result in lower SA, higher workload, and lower search performance as the pilot may spend more than the required time, and resources in processing information for the task. In the context of the current study, ID represents the ratio of relevant information to the total quantity of information. The presentation of task-irrelevant or redundant information on a pilot interface can lead to inefficiencies in performance and increased workload as the pilot will spend their cognitive resources filtering out irrelevant and redundant information. Based on the SEEV Model and Broadbent's (1958) Filter Model, high ID will result in a lower workload and higher SA and search performance. On an eVTOL pilot interface, removing irrelevant and redundant information may lead to reduced workload and more rapid development of SA, which may help improve search performance.

**Review of Past Research on Visual Density and Information density**

Over the past several decades, aircraft pilot interfaces have evolved from traditional analog displays to more technologically advanced glass cockpit displays. One of the major differences between the traditional pilot interface and the pilot interfaces being conceptualized for eVTOL aircraft is the level of clutter associated with the information presented to the pilot. Alexander et al. (2008), based on the characteristics of a display, identified the distinction between the data-driven property of a display (i.e., visual/display density, physical appearance, eccentricity, proximity) and knowledge-driven property of the display (i.e., ID, relevancy, and redundancy) as two widely studied contributors to clutter. This section will provide a comprehensive literature review related to VD, the data-driven aspect of clutter, and ID, the knowledge-driven aspect of clutter. These constructs are proposed to influence pilot performance in the shift from traditional pilot interfaces to simplified eVTOL pilot interfaces. As such, evaluating the impacts of these constructs on pilot SA, workload, and performance was the focus of the current study.

In the literature review that follows, I distinguish the studies related to the two constructs as follows: the VD studies reviewed include those that manipulated the number of display elements per unit of space by either increasing the number of elements present on a display or the spatial proximity of display elements. The reviewed ID studies include those that manipulated the amount of relevant information per total information present on the display by presenting irrelevant and/or redundant information on a display. It should be noted that these are not totally orthogonal constructs, as a manipulation of ID can influence VD, showing some level of interaction between the two constructs. For example, relevant information can be added to a point that the displays become visually dense. The studies reviewed in this chapter are categorized based on the construct they most closely align with. The literature review will lay the foundation for an empirical evaluation of the impacts of VD and ID of simplified, eVTOL pilot interfaces on pilot SA, workload, and search performance.

### *Visual Density*

In research, VD is one of the most frequently adopted and accepted views of clutter and is defined as a surplus of items or objects within a display (Horrey & Wickens, 2004). In the context of the current study, a display is considered visually dense when multiple elements are presented in close spatial proximity or are overlaid over one another (Choe et al., 2021). This often leads to decreased performance because the operator spends additional cognitive resources trying to find the needed information placed among other information in close spatial proximity, which can affect their SA (Alexander & Wickens, 2005; Andre & Wickens, 1989). Several research studies related to the VD aspect of clutter have suggested that high levels of VD in a display can lower the operator's monitoring performance (Schons & Wickens, 1993), increase workload (Ewing et al., 2006; Westerbeek & Maes, 2011), and negatively affect SA (Kim & Kaber, 2009) compared to displays with lower VD. However, research has also shown that higher levels of VD can better facilitate performance, reduce workload, and improve SA compared to low VD displays (Iani & Wickens, 2004). These results suggest that the impact of VD on an operator's SA, workload, and search performance is highly context-dependent.

**Basic Visual Density Research.** A great deal of basic research has been conducted to investigate the effect of VD on visual search performance. These studies have found that high VD display leads to degraded performance. Van de Weijgert et al. (2019) conducted a study to investigate the influence of VD on search performance, as measured by reaction time, by varying the quantity and proximity of information presented on a visual display. For the experimental manipulation, the low VD condition was set up with less information that was sparsely placed on a visual display. In the high VD condition, more information was added to the display, resulting in a "crowded" display, where all the presented information was densely placed. For the experimental task, 12 participants were asked to search for the target present on the display using the low and high VD display conditions. On the display, a red horizontal line was used as the target visual stimuli, which contained a slightly offset gap towards the left or right. The distractor was placed in a vertical orientation. The VD aspect of the visual stimuli was manipulated by increasing the number of targets, the proximity, and the number of distractors on the display. A repeated-measures Analysis of Variance (ANOVA) yielded a significant density effect, showing that the participants' search time was slower with the high VD display condition compared to the low VD condition, $F(1, 11) = 63.68$, $p < .001$, $\eta_p^2 = .853$. The results from this study highlight that increasing VD by adding more information can lead to an increase in search time, resulting in lower performance, as the operator will need time to identify the target information or critical information from densely presented display.

Bennett et al. (2021) conducted a study to assess the effect of the quantity of visual information on search performance in a simulated environment, where 35 participants were instructed to identify a target from the presented information on a visual display. In the experiment, VD was manipulated by increasing the quantity of information on the display. For the experiment, a simulated hallway environment was set up in which participants were tasked to identify the target, i.e., the principal from the crowd, which served as the distractor. In the low VD condition, the number of people in the crowd presented was limited, many of the subjects were sparsely placed. In the high VD condition, the quantity in the crowd was increased, resulting in a densely packed visual display. The high VD condition corresponded to a higher number of, and more closely placed, people in the crowd. Bennett et al. (2021) analyzed reaction time to measure search performance, cognitive processing, and visual search ability. An ANOVA of the reaction time as a function of VD showed that reaction time was quicker for low VD (2212 ms) and slower for high VD (2545 ms). However, the difference in the reaction time was not statistically significant, $F(2, 34) = 1.51$, $p = .220$, $\eta_p^2 = .003$. The results from the study, although not significantly different, suggest a trend that a visual stimulus with high VD can lead to an increase in search time, resulting in lower search performance, primarily because the operator will spend additional mental effort and time locating information potentially resulting in lower performance.

Moacdieh and Sarter (2017) conducted a study to investigate the influence of VD on visual search performance using a simulated graphics program. In the experimental study, VD was manipulated by varying the number of icons on the visual display housing the graphics program. The high VD display used for the experimental manipulation included a complete set of 119 icons, whereas the medium VD condition corresponded to 45 icons. The low VD display condition only had 38 icons. A total of 20 participants were tasked to identify target icons from the presented quantity of icons from each display condition. Visual search performance was measured by calculating the time taken to identify the target icon and analyzed using a three-way repeated measure ANOVA. Only correct answers were considered for calculating the reaction time from the participants' run. The results from ANOVA showed that participants took significantly longer to react with high VD display (9.9 seconds) compared to low VD condition (4.7 seconds), $F(1, 19) = 88.2$, $p < .001$, $\eta_p^2 = .82$. The results from the study suggest that as the quantity of displayed information increases, it takes more time for the operator to locate target information, which leads to poor search performance.

**Aviation-Related Visual Density Research.** Similar to basic research that has investigated the influence of VD on search performance, extensive research has been conducted in the aviation context, specifically to evaluate the influence of pilot interface VD on pilot search performance and SA. Most of the studies have shown that an increase in VD impedes performance and degrades the pilot's SA.

Backs and Walrath (1992) conducted an experimental study to investigate the effect of the VD of an aircraft pilot interface on pilot search time. In the study, VD was manipulated by increasing the number of symbols displayed on the screen. For the low VD condition, 10 sparsely placed symbols were shown to the participants, and in the high VD display condition, 20 densely placed symbols were presented to the participants. As part of the experimental procedure, eight participants were tasked to search for target information on a simulated aircraft tactical display. Participants completed a total of 32 different trials in a single session, in which they were instructed to identify a specific target within six seconds. Participant search time was measured by the time required to locate the target successfully. A within-subject ANOVA revealed that pilot's search time with high VD display condition significantly increased compared to low VD display condition, $F = 17.97$, $MSe = .62$, $p < .001$. Additionally, response accuracy, as measured by participants identifying the correct target, also revealed a significant difference: response accuracy was better in low VD (90%) compared to high VD (78%), $Q(1) = 2.56$, $p < .05$. The authors argued that the participants required more search time due to the increased presence of display elements and showed lower accuracy in the high VD display condition. The results of this study support the current study's hypothesis that as VD increases, by increasing the quantity of information presented on an eVTOL pilot interface, there will be a decrease in in-flight performance.

Wickens et al. (2005) conducted an experimental study to examine the role of clutter in visual search on a pilot interface. The VD aspect of clutter was manipulated to investigate its effect on search performance and traffic detection. In the study, 16 pilots were tasked to search for a target from an array of aircraft on an air traffic display. The array consisted of multiple aircraft that were presented against a black background. VD was manipulated by increasing the array load, i.e., by increasing the number of aircraft on the traffic display. At any given time, there were either three (low VD), five (medium VD), or seven (high VD) aircraft present on the pilot interface. The study's primary objective involved searching for a target aircraft within a pilot interface from the available air traffic information. A within-subject experimental design was used, with two independent variables being manipulated: target presence and the number of aircraft array loads. The participants were required to conduct the target search across 60 trials. The order of trials with different numbers of aircraft was randomized across participants to control for order effects. The results showed that pilots took significantly longer to respond with an increased number of aircraft on the display, $F(2, 30) = 42.71$, $p < .01$. The results from the study demonstrate that as the level of VD increases, there is a significant cost to response time, which Wickens et al. (2005) attributed to pilots spending more time locating information. In the context of the current study, the results support the hypothesis that increasing the VD by increasing the quantity of information on an eVTOL pilot interface can result in lower search performance.

Wickens et al. (2005) conducted another study similar to the experimental study discussed in the previous section. The primary difference between the two studies was that in addition to manipulating the quantity of information, i.e., the number of aircraft, the second study also manipulated the map type by varying the spatial proximity of the navigation information on which the aircraft was presented. In the low VD condition, the display contained sparse features such that the navigation information was placed at a greater spatial proximity from one another. The medium VD condition presented aeronautical chart information. In addition to the aeronautical chart information, the high VD condition also presented terrain information that was placed in close proximity. The manipulation of the number of aircraft and the map resulted in a fully factorial within-subject design. For the tasks, 16 pilots performed similar tasks requiring them to search for a target from the display conditions. Using the same statistical procedure, the results for response time revealed a significant main effect for the map type, $F(2, 30) = 3.91$, $p < .05$, and array size, $F(2, 30) = 3.91$, $p < .05$, showing that participants had significantly higher response time with the high VD condition compared to the low VD condition. When the VD was high, there was a systematic increase in visual search time. These findings support the current study's hypothesis that an increase in VD due to the number and spatial proximity of information can result in lower search performance as measured by response time.

In a study to understand the influence of increasing the quantity and proximity of information on pilot performance, Beck et al. (2012) examined the effects of global and local clutter on aeronautical charts on visual search performance in a piloting task. VD, as a contributor to clutter, was manipulated by using three versions of the same aeronautical chart and increasing the quantity of information and the proximity of the added information in the aeronautical charts. The low VD condition consisted of limited, sparsely placed markers on the aeronautical chart. The medium VD condition added additional information to the existing information, resulting in increased number of markers and closer proximity compared to the low VD condition. In the high VD condition, more information was added to the medium VD display condition, resulting in more information presented in even closer proximity. For the task, three different versions of each base display condition were created, where the target was either placed in a region of high and low VD or the target was completely absent. In the study, 32 non-pilots and 31 F/A-18 pilots in the U.S. Navy were tasked to search for an elevation marker using the aeronautical chart. The participants were presented with 72 charts (24 each for low, medium, and high VD) and were asked to determine whether a target was present or absent in each chart. Responses were coded as accurate (a response was given and it was correct), inaccurate (a response was given and it was inaccurate), or timeouts (no response was given within the one-minute time limit). A 3 x 3 x 2 mixed model ANOVA was conducted with VD (low, medium, high) and target presence (high, low, absent) as within-subjects factors and expertise (pilots, non-pilots) as a between-subjects factor. Results for accuracy showed that there were main effects for global clutter $F(2, 232) = 91.84$, MS = 1.9, $p < .001$, target presence, $F(2, 232) = 155.26$, MS = 4.7, $p < .001$, and expertise, $F(1, 58) = 12.54$, MS = 1.0, $p = .001$ on search performance. Results from the study indicate that the participants were significantly faster at identifying the target with the low VD condition than with the high VD condition. Participant search performance was also significantly lower when VD was high $F(2, 204) = 104.42$, MS = 3350.9, $p < .001$. The results of this study support the current study's hypothesis that as the level of VD increases on a pilot interface, it can impose significant costs on the pilot's ability to locate flight information, leading to lower search performance.

Camacho et al. (1990) conducted two studies to evaluate the effect that the amount of information on pilot interfaces has on performance. The quantity of information as a measure of VD was manipulated by increasing the number of aircraft status indicators on the display. The status display was simulated using a touchscreen cathode ray tube (CRT) display. A total of 24 participants took part in the study, out of which 12 were either current or former military pilots and 12 were non-pilots. The VD of the display was manipulated by increasing the number of indicators on the status display. The low VD display condition consisted of 4 indicators, the medium visual display density condition consisted of 12 indicators, and the high visual display density consisted of 20 indicators. The dependent measures for the study included tracking performance, reaction time, and selection error. For the first experiment, 12 participants were assigned to the group subjected to display using monochrome icons, while the second experiment implemented colored icons on the display. For both experiments, the same format was used for displaying alphanumeric indicators. Both experiments followed the same experimental procedure in which the participants performed a primary tracking task under all experimental conditions. As part of the procedure, participants were provided with a questionnaire, the answers to which were provided on the display. The participants were instructed to track the moving target. An appropriate number of indicators (4, 12, and 20) appeared on the screen during this. The participant's task for both experiments was to keep tracking and selecting the correct answers from the number of indicators on the display. A four-way ANOVA showed that the participant's reaction time significantly increased with an increase in the number of indicators, $F(2, 2856) = 216.24$, $p < .001$. An ANOVA for tracking performance showed a significant main effect for VD, $F(1, 2856) = 10.77$, $p$ < .00l, indicating that the tracking performance decreased as the VD increased. The findings from the study support the hypothesis that increasing VD can negatively influence the pilot's reaction time and search performance.

Alexander and Wickens (2005) conducted a study to evaluate the influence of adding additional weather and surrounding traffic information on a pilot's flight path performance and change detection performance (SA). The pilot interfaces designed for the study were configured in two ways: 2D co-planar and spilt screen, in which VD was manipulated by varying the number of air traffic and weather information visible on the pilot interface. The high VD condition was characterized by a higher number of traffic aircraft and weather icons on the display compared to the low VD display condition. A within-subjects manipulation of the displays was used, in which 24 pilots were tasked to fly with two levels of workload for a total of four conditions presented in counterbalanced order. The pilot's flightpath performance was measured by measuring altitude, lateral deviation, and change detection by measuring their reaction time. Results of a repeated measures ANOVA revealed that pilots had significantly less deviation under low VD conditions, $F(1, 23) = 32.6$, $p < .01$, compared to high VD condition. Change detection analysis revealed that pilots were faster at detecting changes in low VD condition compared to high VD display configuration, $F(1, 23) = 3.20$, $p < .05$. Alexander and Wickens (2005) argued that placing more information induces increased scanning demands, resulting in higher workload, lower performance, and SA. The study's findings support the current study's hypothesis that performance is better when VD on their pilot interface is low.

Although the majority of research has shown that high VD impedes performance, some studies have also shown that an increase in VD can improve performance, and this is often due to the associated increase in ID. Wickens et al. (2004) conducted a study that examined the VD effects of clutter by manipulating the location of the instrument panel, which was either overlaid on an synthetic vision system (SVS) display or located on a display to the side of the SVS. The study's primary purpose was to evaluate the influence of overlaid information on flight performance and traffic awareness using SVS in response to off-normal events. In the study, 14 pilots flew a simulated aircraft with an SVS display through a high-fidelity flight simulator in a terrain- and traffic-rich environment. The proximity of the information presented on a pilot interface was manipulated by varying the spatial position of the instrument panel. In the high VD condition, the instrument panel was overlaid on the SVS display with a tunnel-in-the-sky flight path guidance. In the low VD condition, the instrument panel was presented separately from the SVS display without the tunnel-in-the-sky flight path guidance, resulting in a higher spatial proximity between the presented information. However, it should also be noted that this also increased the ID of the display. As accurately as possible, the pilots were tasked to follow the flight path guidance, verbally report new traffic, and report changes to traffic altitude on the display. Results for flight performance as measured by flight path deviations showed that pilots had significantly lower flight path deviation when using the high VD condition for both vertical deviation, $F(1, 13) = 32.4$, $p < .01$, and lateral deviation, $F(1, 13) = 96.5$, $p < .01$). This is likely due to the increased ID created by adding additional relevant information to the display. In terms of traffic awareness, as measured by other traffic surveillance, results showed the high VD condition imposed a significant six-second cost to detecting traffic compared to when the information was presented on a separate display in the instrument panel, $F(1, 13) = 34.9$, $p < .01$, a finding the authors attributed to clutter. Five of six pilots (83%) who experienced this event in the low VD condition responded with an appropriate evasive response for off-normal event detection. In contrast, only four of eight

(50%) did so in the high VD. Although the difference between these was not statistically significant (Chi-squared = 2.67, $p$ = .102), these results have practical significance given the potentially catastrophic nature of failing to detect conflicting traffic. The findings from the study show that high VD display condition helped improve flight performance. However, this is likely due to the increased relevant information. The findings also showed that there were detriments to event detection due to increased levels of clutter. Wickens et al. (2004) attributed this finding to increased VD, which could lead to lower performance. These results support the hypothesis that there are costs to SA with increased VD of the display. This also provides support for the hypothesis that there is an interaction between VD and ID, such that if increased VD is due to increased ID, it can result in performance improvements.

Iani and Wickens (2004) conducted a study to compare the influence of a traditional, baseline SVS pilot interface with a tunnel-enabled SVS pilot interface on pilot performance. VD was manipulated by using two display configurations. The first display condition, corresponding to high levels of VD, had the SVS terrain overlaid with the instrument panel with a tunnel-in-the-sky flight path guidance. This resulted in a display with a higher number and densely placed information. The second display condition, corresponding to a low level of VD, removed the tunnel-in-the-sky flight path guidance information, resulting in a display with a lower quantity of information that was sparsely placed. The experiment was conducted in a high-fidelity flight simulator with 40 certified pilots. The pilots were required to manually fly three 8-minute curved approaches to land at a synthesized airport over rugged terrain using a digitally depicted environment under instrument meteorological conditions (IMC). An ANOVA was performed on mean absolute flight path deviation data with display format (tunnel vs. baseline) as a between-subjects factor to measure pilot performance. Because the data for performance were not normally distributed, the Kruskal-Wallis test was used to compare the difference between the two display configurations. The results from data analysis showed that the tunnel display ($M$ = 14.8 m, $SD$ = 21.9) supported better flight performance, H = 28.98, $p < .001$, compared to the baseline display ($M$ = 202.6 m, $SD$ = 137.7). This is likely due to the increased quantity of relevant information, which not only increased VD but also increased ID. The results from the study provide support for the hypothesis that there will be an interaction between VD and ID, specifically, if the VD is varied by presenting information that is relevant for the pilots to accomplish the task at hand, thereby increasing ID, which literature has shown can support better flight performance, as discussed in the next session.

***Information Density***

ID is a knowledge-driven property of a display and is the ratio of the amount of the relevant information on a display to the total amount of information presented; and is associated with the information relevancy and redundancy for the task at hand (Doyon-Poulin et al., 2014). In the context of the current study, a pilot interface with a greater number of relevant pieces of information in the absence of redundant or irrelevant information is categorized as a high ID display condition. A pilot interface with fewer relevant pieces of information and an increased number of irrelevant or redundant information is categorized as a low ID display condition. Research has shown that pilot interfaces with high ID improve performance and SA and lower workload (Alexander et al., 2003; Brahydt & Hansman, 1990; Morphew & Wickens, 1998). While the presence of more relevant information has been shown to aid pilots in improving their performance, research has also shown that high levels of ID, even when the information is relevant, can lead to high VD, which can increase workload and be detrimental to pilot performance (Lohrenz & Hansman, 2004).

Barhydt and Hansman (1999) conducted an experiment to evaluate the influence of providing additional aircraft intent information on a prototype cockpit traffic display. In the study, ID was manipulated by systematically increasing the relevant information related to the intruder aircraft intent information on the prototype cockpit traffic display. The experiment's primary objective was to examine the effect of increasing levels of traffic information on the pilot's ability to recognize separation violation (SA) and maneuvering time (performance). Each display was superimposed onto a traditional map display with current traffic collision avoidance system (TCAS) symbology and aircraft ownship identity. Using a within-subject design, the ID of the cockpit traffic display was manipulated between four separate displays (baseline TCAS, rate, command state, and flight management system path (FMS-path) with increasing amounts of relevant intruder traffic information, respectively. The baseline TCAS display condition corresponding to the low ID condition only displayed the intruder aircraft's identity, climb rate, position, and relative altitude to the ownship aircraft. The rate TCAS display, in addition to the information presented in the baseline display condition, added a conflict probe and profile view in relation to the intruder aircraft. The command state display condition included the intruder aircraft's commanded heading and altitude in addition to the information shown on the rate display. The FMS path added the intruder aircraft's programmed trajectory, lateral navigation (LNAV), and vertical navigation (VNAV) path information to the command state display condition. The baseline TCAS display corresponded to low ID conditions with four pieces of intruder aircraft information. While the rate, command state, and FMS path corresponded to the high ID condition with relevant intruder aircraft information, respectively. The experiment was carried out using a Massachusetts Institute of Technology (MIT) part-task flight simulator, in which eight commercially rated pilots were tasked to maintain a minimum of 2 nm lateral or 500 ft vertical separation from all other traffic. McNemar tests for correlated proportions, with a one-tailed *t* test, were performed to compare separation violation percentages between the low and high ID display conditions. Performance data analysis showed that compared to the baseline TCAS display condition, the pilot's tendency to violate separation from the intruder aircraft was lower with rate, command state, and FMS-path display conditions. However, the difference was not significant at the 95% confidence interval (TCAS and rate, $t(47) = .24$, $p > .05$; TCAS and command, $t(47) = 1.17$, $p > .05$; and TCAS and FMS-path, $t(47) =$

1.71, $p > .05$. Even though the results were not statistically significant, Barhydt and Hansman (1999) argued that due to the presence of relevant information, the participants were better able to avoid the intruder aircraft and maintain separation as the relevant information assisted the pilots in maintaining their separation from the intruder aircraft. As measured by pilots performing an avoidance maneuver, results for maneuvering time showed that pilots performed significantly better with high ID display conditions than with low ID conditions. The maneuver time differences for the following display combinations were statistically significant at the 95% confidence interval ($p < .05$), TCAS and rate, $t(26) = 2.92$, $p = .043$; TCAS and FMS-path, $t(26) = 5.66$, $p < .001$; rate and FMS-path, $t(26) = 3.85$, $p = .004$; and commanded state and FMS-path, $t(26) = 4.01$, $p = .003$. In the context of the current study, the results from Barhydt and Hansman (1999) support the hypothesis that increased ID due to the presence of more relevant information on a display will help improve pilot flight performance.

Morphew and Wickens (1998) examined the effect of different information-varying displays on pilot performance, workload, and traffic avoidance (SA). In the study, three pilot interfaces were conceptualized with varying levels of ID. The low ID condition (the baseline display) only provided a predictive flight path of the ownship. The medium ID condition (the intruder predictor display) provided both the ownship and the intruder aircraft flight paths. While the high ID condition (threat vector display), in addition to both aircraft having flight path predictor information, also displayed the potential flight path direction towards the point of closest contact with the intruder aircraft, which was relevant for the pilots to complete the experimental task. The study was simulated in a low-fidelity flight simulation, where 15 pilots were instructed to fly to a designated waypoint while avoiding traffic conflicts and minimizing the deviations from the established speed, heading, and altitude. Each simulated run involved 10 encounters with a single intruder aircraft. Pilots were tasked to respond to the intruder as quickly as possible. The results for traffic avoidance, as a measure for SA, revealed a significant difference between display types, $F(2, 28) = 19.28$, $p < .001$, demonstrating that the pilots showed better traffic avoidance with high ID compared to low ID display condition, as they had more relevant information available to make evasive maneuvers to avoid conflict with the intruder aircraft. Authors report that analysis of workload by the NASA TLX score revealed a marginally significant difference between the display types, $F(2, 26) = 3.08$, $p = .08$, with the threat vector yielding the lowest workload scores. The result from the study shows that high levels of ID support better performance, which Morphew and Wickens (1998) attributed to task-relevant flight information. In the context of the current study, the results support the hypothesis that increased ID as a result of increased relevant information can result in lower mental demand, resulting in reduced workload and improved SA.

Alexander et al. (2003) conducted an experiment to evaluate the influence of display format on pilot flight performance, traffic awareness, and workload. In the study, 18 pilots were tasked to fly a sequence of six flight scenarios designed to compare three levels of information in the pilot interface. All display formats were overlaid on a computer-generated terrain. ID was manipulated by removing relevant information from the display conditions. In the high ID display condition, relevant flight path guidance information in the form of a highlighted tunnel-in-the-sky display was presented over the terrain information. In the medium ID display condition, relevant flight path guidance information in the form of tunnel-in-the-sky was presented. However, it was not displayed as prominently in the first display. The low ID display condition presented no relevant flight path guidance information. Instead, it displayed the aircraft's position five seconds ahead of ownship. A within-subject, counterbalanced manipulation of display type was used such that each pilot experienced each level of ID once and again in reverse order. Flight performance was measured by calculating the vertical and lateral flight path deviations during the flight scenarios. A repeated measures ANOVA revealed that participants had significantly lower vertical flight path deviations with medium and high ID display conditions than low information display condition, $F(2, 34) = 6.05$, $p < .01$. Lateral deviations were measured as deviations from the center of the path. Using planned comparisons analysis revealed that pilots showed fewer lateral deviations with medium ID ($M = 7.18$ m) than with high ID ($M = 7.74$), $t(17) = -1.8$, $p < .09$, and smaller deviations with high ID than the low ID, $t(17) = -2.6$, $p < .02$. Traffic awareness, as measured by traffic detection, showed that the pilots were not significantly slower in identifying traffic with high ID ($M = 15.5$ s) compared to medium ID ($M = 11.1$ s) and low ID ($M = 10.7$ s), $F(2, 32) = 2.37$, $p =$

.11. The subjective mental workload was highest with low ID, intermediate with medium ID, and lowest for high ID display condition. However, the study did not discuss the *p* values for the workload results.

Prinzel et al. (2018) conducted a study to evaluate the influence of using an SVS display on pilot workload and SA. ID was varied by providing additional relevant information on the pilot interface. The lower ID display condition consisted of a traditional blue-over-brown PFD without relevant terrain elevation information. The higher ID display consisted of an SVS display with added relevant terrain elevation information. The study's primary objective was to determine whether including relevant terrain information on a pilot interface influenced the pilot's ability to maintain altitude and a leveled flight. In a high-fidelity B-787 simulator, pilots were tasked to fly a pre-determined flight path and were asked to maintain altitude and airspeed. For the low and high ID display condition, the pilot's workload as measured with the NASA-TLX showed a significant main effect, $F(1, 11) = 8.952$, $p < .012$, showing that the pilot using SVS display reported lower workload compared to blue-over-brown display. The results for SA, as measured by Situation Awareness Rating Techniques (SART), showed that the pilot using the SVS display gave it a significantly higher SA rating compared to the traditional blue-over-brown display, $F(1, 11) = 9.329$, $p < .001$. Prinzel et al. (2018) state that the presence of the relevant terrain information in the SVS display increased the pilot's confidence to maintain terrain clearance, thereby increasing their SA and reducing workload, which was not possible with the blue-over-brown display. The results from the study align with the current study's hypothesis that increased ID through the presence of relevant information will increase pilot SA and lower their workload.

Lazaro et al. (2021) investigated the influence of visual complexity on visual search and target detection on cockpit displays. In the study, what they refer to as visual complexity and what I refer to as ID of a pilot interface was manipulated by increasing the level of irrelevant information presented on the display. From the visual stimuli created as part of the experimental manipulation, the low visual complexity display corresponded to a high ID condition as it only had relevant information about the simulated task with minimal irrelevant information. The medium visual complexity corresponded to the medium ID condition, which had more irrelevant information than the high ID condition. The high visual complexity display condition corresponded to the low ID display condition, which consisted of more irrelevant information than low and medium ID conditions. The simulated experiment was conducted using a 23-inch monitor replicating an F-35 cockpit display, for which 17 participants were presented with varying levels of ID. The participants were tasked to identify whether the target was present or not in each trial. Two task performance measures (response time and response accuracy) were assessed. A one-way repeated-measures ANOVA was conducted at a significance level of .05 to investigate the effects of ID on performance. Results for task performance as measured by the number of correct responses revealed a significant effect of ID on response time, $F(2, 32) =$

198.07, $p < .001$, demonstrating that participants were slower in detecting targets using high ID compared to medium and low ID condition. The results for response accuracy as measured by participants correctly identifying the targets showed that participants were significantly more accurate in identifying the target with high ID condition than with low ID condition ($\chi 2(2) = 18.72$, $p < .001$). The results support the hypothesis of the current study that decreasing ID by including irrelevant information on the displays can lead to lower performance.

Moacdieh et al. (2013) conducted a study to investigate the performance and attentional costs associated with PFD clutter. A simulated flight was conducted with 23 instrument-rated pilots (22 males and one female), of which nine held an airline transport pilot (ATP) rating, seven were commercial pilots, and five were private pilots. Three generic PFD designs (low-, medium-, and high-cluttered) were created for the study. In the context of the current study, the low clutter display condition corresponded to the high ID condition as it had more relevant information than irrelevant information. The medium clutter condition corresponds to the medium ID condition, which included a few irrelevant information elements in addition to the relevant information. The high clutter condition in the study corresponded to the low ID due to the increased presence of irrelevant information included to create a cluttered display. A between-subject experimental design was employed for the study in which the pilots were tasked to fly a 32-minute flight from Denver, CO, to Aspen, CO. During the flight, there were two high-workload phases of 8 minutes each and one low workload phase for 16 minutes. Throughout the flight, 22 visual alerts and notifications (11 each in high and low workload) appeared on the PFD. Search performance was evaluated by measuring the reaction time for the pilots to identify the alerts and notifications. A mixed-design ANOVA after applying Bonferroni corrections for post-hoc analyses showed that high ID led to a significant reduction in response time to alerts, $F(2, 163) = 4.47$, $p = .013$, suggesting that pilots were able to react faster using high ID displays compared low and medium ID displays. Data analysis for the high workload phase yielded similar results, showing that the pilot's reaction time was significantly lower with the high ID condition compared to the medium and low ID condition, $F(2, 189) = 4.65$, $p = .011$. Although a similar trend was observed during the low workload phase, the difference was not significant. Moacdieh et al. (2013) concluded that including irrelevant information not critical for the task can lead to the pilot missing critical information such as alerts and notifications. Researchers proposed that this is due to the pilot needing to spend additional cognitive resources filtering out irrelevant information, leading to their search performance detriment. In terms of workload, although the low workload condition did not yield a significant difference, the findings reveal a trend suggesting that under high and low workload phases of flight, the presence of relevant information facilitated better search performance. In the context of the current study, this supports the hypothesis that low ID can result in lower search performance.

Lohrenz and Hansman (2004) conducted a study to investigate the influence of display content and clutter on pilot performance. The level of ID was manipulated in three display conditions. The low ID corresponded to the map-only display with just the topographic information. The medium ID corresponded to a combination map display, which consisted of a topographic map and the flight path, meaning that the information displayed had some level of relevant information present, with moderate presence of irrelevant information. The high ID corresponded to overlays-only display conditions with only mission-specific, relevant flight path information and no irrelevant information. In the study, 12 volunteers flew one mission for each display. Subjects' primary target acquisition task was to stay on the flight path and identify a target as quickly as possible. During the task, participants' flight guidance performance (FP), target acquisition performance (TP), flight guidance workload (FW), and target acquisition workload (TW) were measured. A least square analysis for flight guidance measure identified display condition as having a main effect on performance, $F(2, 20) = 14.71$, $p < .001$, with the participants performing significantly worse with low ID display condition compared to the high ID display condition. Results for workload as measured by the participant's target acquisition using a *t* test revealed a significant difference in TW rating between medium ID and high ID condition, $t = 1.92$, $p < .1$, suggesting that the participants experienced significantly higher workload with low ID display compared to the high ID display condition. The results from the study support the current study's hypotheses that a low level of ID results in lower search performance and higher workload compared to a high ID, which suggests increased search performance and lower workload.

Some research indicates that decreased ID in the form of redundant information can lead to improved search performance and increased workload. Peterson et al. (1999) examined the effectiveness of using a terrain-enhanced PFD (TE-PFD) for detecting and avoiding potential Controlled Flight into Terrain (CFIT) accidents. For the study, two display configurations were set up. While both displays showed terrain information, including flight information, one crucial difference between the two display conditions was including a separate artificial horizon line and the natural horizon line provided by the terrain. In the context of the current study, the display configuration with the artificial horizon line corresponded to a low ID condition, as the artificial horizon line was deemed redundant for the task. The second display configuration without the artificial horizon line corresponded to high ID due to the absence of redundant information. The two displays were simulated on a flat panel CRT monitor, in which 22 participants were tasked to judge potential conflict between the current flight path and surrounding terrain using the two display configurations. Participants viewed the TE-PFD in a variety of terrain situations, either with or without an artificial horizon line superimposed over the terrain. Participants were asked to judge whether an avoidance maneuver was required. In addition to the artificial horizon line on the pilot interface, the horizontal separation (distance) between the aircraft and terrain was manipulated between 2000 m, 4000 m, and 6000 m, and the vertical separation between the aircraft was manipulated between 150 m and 50 m. Reaction time was measured from the display onset to the pilot deciding whether to maintain course or perform a maneuver to avoid the terrain. The study found a significant difference in reaction time between low and high ID display conditions, $F(1, 21) = 11.57$, $p < .01$. With the low ID condition, i.e., when the horizon line was present, the participants' average reaction time was 926 ms. While in the high ID condition, in which the horizon line was absent, the average reaction time was 1778 ms, nearly twice as long. Although the result from the study seems to show that low ID led to lower search performance due to an increase in the clutter from the presence of redundant information, Peterson et al. (1999) argued that the benefits of using redundant information to aid pilot search performance outweighed the cost incurred due to clutter. The results from the study suggest that redundant flight information does not always lead to negative search performance impacts if the information is relevant to the task at hand and can aid the pilot in performing the task at hand.

Van Geel et al. (2020) experimentally evaluated the effect of enhanced vertical flight information on the pilot's workload and search performance. The study's manipulation consisted of two experimental pilot interfaces. The baseline display condition corresponding to high ID consisted of a basic PFD and a vertical situation indicator (VSD) with only relevant information for the experimental procedure. The Configured VSD (CVSD) and the PFD corresponded to low ID display conditions as they displayed several irrelevant pieces of information that were not required for the flight procedure. A simulated pilot-in-the-loop experiment was carried out with 16 pilots. All participants flew two similar scenarios with both display conditions in a fixed base flight simulator. Participants were tasked to fly an approach and go-around procedure flight maneuvers in a single scenario. The scenario consisted of a standard non-precision approach using a 3-degree glideslope with a go-around at an altitude of 600 ft. The experiment was set up as a within-subjects repeated measures design, meaning all participants flew using the CVSD and baseline VSD displays. The dependent measures for altitude and velocity were measured using Root Mean Squared Deviation (RMSD) to assess pilot performance. Workload was measured using the Rating Scale Mental Effort (RSME), and SA was measured by evaluating the pilot's awareness of their casual and intentional velocity limits. Results showed no significant difference in altitude performance between the two displays using the Wilcoxon-Signed-Rank test, $Z = .362$, $p > .05$. Velocity RMSDs were compared using paired *t*-test, which showed that pilots had significantly better velocity tracking performance with the high ID display condition (baseline VSD), $t(15) = –2.19$, $p < .05$ compared to the low ID display condition (CVSD). Data analysis for SA using the Wilcoxon-Signed-Rank test showed that pilots showed significantly better awareness of their current state with the low ID display condition, $Z = –2.29$, $p < .05$, compared to the high ID condition. The workload was analyzed objectively through control input variation and subjectively through self-reported RSME ratings. Control inputs were captured by measuring side stick deflection and thrust setting. For both display conditions, Wilcoxon-Signed-Rank tests found no difference for standard deviation in elevator input rates, $Z = –.958$, $p > .05$, or throttle deflection rates ($Z = –.675$, $p > .05$).

These results are replicated by self-reported workload ratings, showing no differences in reported subjective workload scores, $t(15) = -.798$, $p > .05$. Van Geel et al. (2020) argued that with low ID display conditions with additional information, the pilots were able to maintain an adequate level of safety by abiding by the flight envelope information, even though it was not critical to the task they were performing compared to when using the high information display condition. The findings from the study highlight that the pilots might place a higher emphasis on maintaining safety rather than optimizing performance. In the context of the current study, the findings suggest a clear interaction between visual and ID aspects of a visual display. While the high ID did support better search performance, an equivocal result for SA showed that low ID did support the pilot maintaining better SA.

**Interaction Between Visual and Information density**

The previous two sections covered the literature on VD and ID as two separate constructs. This section will provide a review of research conducted to understand the impact of the interaction between VD and ID on pilot SA, workload, and search performance. Research has shown that high ID is good as long as it does not lead to VD that is too high. However, there appears to be a point at which too much relevant information is presented, resulting in VD that is too high and negatively impacts search performance. Therefore, high ID leads to improved search performance, reduced workload, and increased SA when VD is low or manageable. However, when ID is too high, it can lead to high VD, leading to search performance degradation, increased workload, and decreased SA.

Doyon-Poulin et al. (2014) investigated the influence of three experimental PFDs with varying levels of VD and ID aspects of clutter with similar flight guide functions (localizer and glideslope instruments) on pilot workload and performance. In the low clutter display, VD was low, and ID was high with basic readouts of the relevant flight information. In the medium clutter display condition, Doyon-Poulin et al. (2014) increased ID, which also increased the VD to a medium level but was still manageable. In the high cluttered display, the authors increased the ID to the point that the VD of the display was very high. For the study, 12 pilots were tasked to complete a total of nine simulated approaches to Montreal Airport's runway 06L in a fixed-base, side stick-controlled flight simulator. The pilots manually flew all the approaches with the help of localizer and glideslope instruments in instrument meteorological conditions (IMC). For the scenario, the pilots were instructed to keep the current heading and follow the localizer and glideslope indications as precisely as possible. The pilots completed the approach using each of the three PFDs, employing a within-subject experimental design. The pilot's mental workload was measured using the NASA-TLX scale, and performance was assessed in terms of localizer and glideslope deviations. Results from the study indicated that there was a significant effect of PFD clutter on pilot workload $F(2, 22) = 8.67$, $p < .005$, partial–$\eta^2 = .44$. Tukey's HSD post-hoc test revealed that participants' workload scores significantly increased with increase in level of clutter, such that pilots exhibited higher workload with the high clutter condition compared to medium and low clutter display conditions. These results suggest that when the presence of relevant information is too high, it makes the display visually dense compared to the low and medium clutter. When the ID was high, but the VD was manageable, the pilots were able to perform the task without increasing their workload. Regarding flight performance, only difference in localizer deviations between display types reached statistical significance, $F(2, 22) = 3.70$, $p < .05$, partial–$\eta^2 = .25$. Tukey HSD post-hoc test revealed that the pilots performed better with medium clutter display condition compared to high clutter display configuration. The performance results also reveal a similar trend to the workload results, demonstrating that when a display is visually dense due to high ID, it can cause performance decrements. The workload and flight performance results suggest an interaction between the VD and ID aspects of a display. When the ID is high with low VD, it can support better performance and lower the pilot's workload. However, once the threshold is reached in which high ID results in a visually dense display, it can cause decrements to the pilot performance and increase workload. The study's findings support the hypothesis that there is an interaction between VD and ID. While relevant information has been shown to support pilot performance, there is a limit to how much relevant information can be presented on a pilot interface. Too much relevant information after a certain point can make the display visually dense, with a cost to the pilot's workload and performance.

Alexander et al. (2008, 2009) conducted a study to evaluate the effect of data-driven and knowledge-driven contributors of clutter on pilot workload and performance. For the study, VD and ID as a function of clutter was manipulated by varying the quantity and the spatial proximity of information on the pilot interface. The experiment was conducted using the Integration Flight Deck (IFD) simulator, where four expert pilots were instructed to fly six approaches. The low clutter display corresponded to a low VD and high ID display condition as it presented a comparatively lesser quantity of information, the majority of which was relevant information for the task. In the medium clutter display condition, the authors increased ID, which also increased the VD to a medium level but was still manageable. In the high cluttered display, the authors increased the ID to the point that the VD of the display was very high. The display configuration and workload were manipulated as within-subject variables such that each participant got the chance to fly with low, medium, and high clutter display configurations under both low and high workload conditions. Participants' performance was assessed by measuring glideslope and localizer deviation, while mental workload was measured using NASA-TLX. The analysis of workload data showed a significant effect of display configuration on workload, $F(2, 93) = 3.3$, $p = .04$. This suggests that participants experienced reduced workload when the display had greater ID while maintaining a manageable VD level. In terms of performance, data analysis revealed that display configuration had a marginally significant effect on glideslope deviation, $F(2, 59) = 2.75$, $p = .07$. The medium clutter yielded the most stable control compared to the high clutter display configurations. The workload result from the study showed that high clutter displays produced elevated reports of workload due to increased VD of information. In comparison, the low clutter displays were also reported with higher workload scores because of spaced-out information required for experimental tasks, meaning that the pilot would exert additional effort to locate information. In the context of the current study, although added display elements provide pilots with relevant information, the imposed cost of VD exceeded the benefits of relevant information for a specific task. This supports the current study hypothesis that an interaction between VD and ID could influence pilot performance, SA, and workload.

Alexander et al. (2012) conducted a study to investigate the effect of low-, medium-, and high-clutter on pilot performance. The study was conducted using the IFD in which three levels of clutter were manipulated by adding relevant and irrelevant flight information to the critical flight information. The low clutter condition corresponded to high ID and low VD by sparsely placing limited flight-relevant information. The medium clutter display consisted of additional flight information to the information already present in the low clutter display condition, resulting in a high but manageable VD display condition. In the high clutter condition, the ID of the display was increased by including an SVS image overlaid on the flight information, making the display more visually dense. For the study, six airline captains were required to fly six instrument landing system (ILS) approaches, where the primary task for the pilots was to maintain the localizer course at a constant altitude before intercepting the glideslope. The dependent measures included measuring the pilots' workload using NASA-TLX and performance measured by calculating the deviation from the glideslope and localizer path. A repeated measures ANOVA showed that pilots reported significantly lower workload with medium clutter compared to high clutter display conditions, $F(2, 93) = 3.3$, $p = .04$. A repeated measures ANOVA for flight path tracking performance revealed a marginally significant effect of the clutter level on glideslope deviation, $F(2, 59) = 2.75$, $p = .07$, revealing that the pilots deviated maximum under low-clutter condition. In comparison, medium-clutter condition showed the least deviation compared to high-clutter conditions. In the context of the current study, the results from Alexander et al. (2012) reveal a trend that there is an interaction between high and low levels of VD and ID. While high ID supports better performance and lowers workload, there is an interaction between the two constructs such that after a point, high ID leads to an increase in the VD of the display, which research has shown to impede performance.

**Summary and Study Implications**

In conclusion, in both basic and aviation research paradigms, increasing the VD by increasing the quantity and the spatial proximity of information per space available on a visual display can lead to lower SA and lower search performance. Increased ID, as defined by the ratio of the quantity of relevant information with the total quantity of information presented to the pilot, can lead to improvements in search performance, lower workload, and higher SA. However, under certain situations, added redundancy did provide some level of benefit over clutter as it provided task-relevant information. However, there is a threshold at which this benefit is achieved. When looking at both these constructs together, there is an interaction between VD and ID. When VD is low, high ID leads to higher search performance and lower workload. However, after a threshold is reached, when ID causes an increase in the VD of the display, it can lead to decrements in pilot search performance and increased workload.

While VD and ID's influence has been studied using conventional, traditional pilot interfaces, research using pilot interfaces for an eVTOL aircraft, particularly focusing on AAM operations, has not been conducted. This is a new research domain as there are differences between conventional and AAM flight operations concerning the information displayed on the interface and how pilots will use this information. This research offers new insights built from findings from the current literature with basic and conventional aviation research contexts by examining the impact of VD and ID on pilot SA, workload, and search performance in an AAM eVTOL piloting context. The SEEV model and Broadbent's Filter Model of Attention provide theoretical support, and the literature reviewed provides empirical support for the current study research questions and hypotheses presented in Chapter 1.

# Chapter 3

# Methods

## Population and Sample

### *Population*

The target population for the current study is all United States (U.S.) pilots, who hold an instrument rating and operate either scheduled or non-scheduled air taxi and cargo missions. Recently, the FAA proposed training and pilot certification requirements for air taxi missions in the AAM ecosystem (FAA, 2023c). In these requirements, the FAA expects commercial pilots of winged eVTOL carrying passengers to hold a powered-lift category and instrument rating, in addition to the type certification for each eVTOL they will fly. This will be required in addition to any airplane or helicopter rating a pilot already might have (FAA, 2023c). According to the FAA, the minimum flight time needed to earn a commercial rating under FAA Part 61 is 250 hours (FAA, 2023a), and under FAA Part 141 is 190 hours (FAA, 2023). As eVTOL aircraft are currently not type certified to operate commercially and are not used for flight training, pilots certified to fly eVTOLs are not accessible because they do not exist yet. Considering the requirements identified by the FAA for eVTOL pilot training, the most suitable population is pilots who have an instrument rating, so I recruited from this participant pool.

The reader should keep in consideration that the FAA classifies eVTOL aircraft as a powered-lift aircraft from an aircraft certification point of view (FAA, 2023c). The recently announced pilot training requirement for powered-lift aircraft is not a final version and is made available to the industry stakeholders for comments and inputs; a final version has yet to be published by the FAA. Due to the nascent nature of the AAM operations and the uncertainty in the minimum requirements for crewed eVTOL operations, the minimum flight time required for an eVTOL pilot is subject to change in the future as the FAA releases further regulations in support of crewed powered-lift aircraft operations.

The accessible population included all the student pilots currently enrolled at the Florida Institute of Technology (FIT), who hold their instrument rating. Flight students receive this certificate rating at flight schools like FITA, so I recruited from this participant pool. According to the Bureau of Labor Statistics and the United States (U.S.) Census Bureau, the demographic breakdown of the pilots who hold an instrument rating is presented in Table 3.1 for the target population.

**Table 3.1**

*Target Population Demographics*

| | | | | Race[b] | | | |
|---|---|---|---|---|---|---|---|
| Occupation by Location | Total pilots (2023)[a] | Female | Male | Caucasian | African American | Asian | Hispanic Latino |
| **Instrument Rated pilots** | | | | | | | |
| United States | 483,255 | 16.4% | 83.6% | 81.7% | 8.9% | 1.1% | 7.4% |

*Note.* [a]Reprinted from Federal Aviation Administration Airmen Statistics. (2023).

[b]Reprinted from United States Census Bureau. (2022).

### *Sample*

From the accessible population, a convenience sampling strategy was employed to recruit participants for the current study. As the future cadre of eVTOL pilots could come from Part 141 and Part 61 flight schools, participants who had completed their instrument rating were considered to participate in the study. Demographic information of the sample was collected and compared against the population after data collection to ensure a representative sample. Descriptive statistics of the sample demographic data, presented to the reader for generalizability purposes in Chapter 4 (see Table 4.1), showed a similar ratio of male to female participants when compared to the target population; however, the sample was less diverse than the population with respect to race. The reader should keep this in mind when interpreting the results.

### *Power Analysis*

Based on the lack of previous research conducted in line with the current research, an a priori repeated measures within-factors Multivariate analysis of variance (MANOVA) power analysis was performed using G*Power (Faul et al., 2009) with the following assumed values:

- $\alpha = .05$
- Effect size = .25
- Power = .80
- Number of groups = 4
- Number of measurements = 3

- Correlation among rep. measure = .50

The power analysis resulted in a needed sample of 32 participants, which was the number of participants the study aimed to recruit; however, due to difficulty in recruiting qualified participants to volunteer, a sample of only 26 participants was obtained. However, a post hoc power analysis indicated a power of .86, revealing that the sample of 26 participants was sufficient to find the hypothesized effects.

**Human Subject Research**

An Institutional Review Board (IRB) expedited request was submitted and approved by the Florida Institute of Technology (FIT) IRB (see Appendix A). The risks of participation in this study did not exceed the risks of the everyday operation of a flight simulator on a desktop computer. The primary risk was potential simulation sickness, which was mitigated with breaks as needed by participants and the ability for participants to stop participation at any time. The researcher was the sole data collector of the study. The researcher and the major advisor were the only individuals with access to the participant data. Participant identifying information was kept separate from their data to ensure anonymity.

**Research Methodology**

The current study employed a repeated measures experimental research design. This was the ideal and appropriate method to explore quantitative, experimental data to show statistical interaction between varying levels of VD and ID of an eVTOL pilot interface and capture the pilot’s reaction to an eVTOL pilot interface and its associated level of clutter. The quantitative part of the research methodology was a within-subjects repeated measures design with two independent variables, including VD (low VD vs. high VD) and ID (low ID vs. high ID). This was the appropriate research methodology to explore quantitative data to show statistical differences between the VD and ID levels, any associated interactions, and their influence on the participant's SA, workload, and search performance. Participants in this study had varying levels of experience and came from various backgrounds. To control for individual differences, a within-subject design was used. As such, there was only one group, and each participant received all treatment conditions in a counterbalanced order to avoid order effects. The current study measured three dependent variables: SA, workload, and search performance. The participant's SA was measured using SAGAT queries (Endsley, 1995), workload was measured using the NASA-TLX questionnaire (Hart, 1986), and search performance was the time in seconds taken by the participants to name the final approach fix waypoint upon asking. Utilizing this approach allowed the comparison of varying VD and ID levels on the multiple dependent variables.

For the qualitative part of the study, a phenomenological approach was used as the research question aimed to capture the participant's reaction to flying a simulated eVTOL aircraft with varying levels of pilot interface VD and ID. A complete list of qualitative questions is provided in Appendix B.

**Independent Variables**

In the current study, two independent variables (IVs) were manipulated with two levels each: (1) VD (low vs. high) and (2) ID (low vs. high). The two IVs were manipulated by adding and removing customizable pieces of information from the experimental testbed's pilot interface. Specifically, VD was changed by manipulating the number of pieces of information presented on the displays. Critical information was presented on the pilot interface in all conditions; relevant, irrelevant, and redundant information was determined and then added or removed to develop a total of four display conditions. The second IV, ID, was manipulated by adjusting the ratio of the number of pieces of relevant information presented on the display to the total number of pieces of information presented on the display. This total number of pieces of information is the sum of relevant, redundant, and irrelevant information.

To determine the display conditions for the experimental task, the following process was used:

1. The first step was to determine what information would be presented on an eVTOL pilot interface. To accomplish this, I conducted a comprehensive review to investigate the current generation of eVTOL aircraft design, focusing on the pilot interface. From the review, I identified all the information poised to be presented on an eVTOL pilot interface. A complete list of all the information that is presented on the simulation testbed is included in Appendix C.
2. After determining which pieces of information would be presented on an eVTOL pilot interface, I met with five industry SMEs with extensive experience in the eVTOL domain and asked them to mark which pieces of

information were critical, relevant, irrelevant, or redundant to the experimental task. In the context of the current study, relevant information refers to any piece of information that is useful or is needed for the pilot to complete a particular task at hand. This includes information that would help the pilot assess the current operational status of the flight (FAA, 2014). Critical information refers to any piece of information that is essential for the pilot to complete a task at hand. Irrelevant information refers to any piece of information that is not required or helpful for the pilot to complete a task. Redundant information refers to any information that is presented at multiple locations on the pilot interface. The SMEs included two helicopter pilots, one eVTOL test pilot and aircraft researcher, one eVTOL operation SME, and a retired FAA Acting District Manager of Terminal Facilities.

3. I also queried the SMEs regarding which information could be removed from the pilot interface without impacting their ability to land an eVTOL aircraft. From the input provided by the SMEs, I created a complete list of all the relevant, irrelevant, critical, non-critical, and redundant information, which is provided in Appendix D (PFD) and E (MFD).
4. Inter-rater reliability was calculated as the average percent agreement for each piece of information. Each piece of information was scored 1 if it was tagged relevant and critical information by the SME and 0 if the information was tagged irrelevant and non-critical information by the SME. For each piece of information on the pilot interface, the percentage agreed

upon was calculated. Then, the overall average of these percentages was calculated, yielding an overall inter-reliability percentage of 82.6%.

5. Next, I created quantitative metrics for VD and ID. These metrics were developed based on definitions provided by Alexander et al. (2008) and Moacdieh and Sarter (2014) for VD and ID as measures of clutter. Utilizing the VD and ID metrics, as shown in equations (1) and (2), allowed me to quantify and ensure a quantifiable difference between the two levels of each condition for the two constructs. These metrics were used to determine the total quantity of information and the ratio of relevant information to the total quantity of information. The total quantity of information included relevant, irrelevant, and redundant information.

$$\text{VD} = \text{Total quantity of information} = \text{Quantity of Relevant information} + \text{Quantity of Irrelevant information} + \text{Quantity Redundant Information} \quad (1)$$

$$\text{ID} = \text{Quantity of Relevant Information}/\text{Total Quantity of Information} \quad (2)$$

6. Utilizing equations (1) and (2), I calculated the ID ratio and VD levels for each display condition. I then performed an iterative process, in which I adjusted the levels VD and ID to ensure sufficient difference between the display conditions. Once the four display conditions were set, the SMEs were asked to rate the level of clutter for each display condition.

7. I then performed an iterative process in which I adjusted the information included in each of the four display conditions. I evaluated them objectively using the metrics and subjectively based on the SME clutter ratings of the eVTOL pilot interface until I reached a state where there were distinct

differences between the low and high levels but no distinct differences within each level.

8. After the four display conditions were finalized, I completed a pilot run with five participants in order to examine the trends in the participants' SA, workload, and search performance to validate that the manipulations were distinct enough. Based on the results, appropriate adjustments were made to the four display conditions. The final SME ratings are shown in Table 3.2, and the values used in the calculations and the resulting VD and ID levels for each display condition are shown in Table 3.3.

**Table 3.2**

*SME Subjective Rating of Clutter for Display Conditions*

| | Low VD, low ID | Low VD, high ID | High VD, high ID | High VD, Low ID |
|---|---|---|---|---|
| SME 1 | 3 | 3 | 4 | 5 |
| SME 2 | 2 | 3 | 4 | 5 |
| SME 3 | 2 | 3 | 4 | 5 |
| SME 4 | 3 | 2 | 4 | 5 |
| SME 5 | 4 | 4 | 5 | 5 |
| *M*[a] | 3 | 3 | 4 | 5 |

*Note*. 1 = Not cluttered at all, 2 = Slightly cluttered, 3 = Moderately cluttered, 4 = Cluttered, 5 = Extremely cluttered

[a]*M* = Mean for each display condition rounded to the next whole number

**Table 3.3**

*Display Condition Calculations*

| Ratio Calculation | Low VD, high ID | High VD, high ID | High VD, low ID | Low VD, low ID |
|---|---|---|---|---|
| $R_{critical}$ | 18 | 21 | 21 | 18 |
| Total quantity of relevant information ($R_1 = R_{critical} + R_{non\text{-}critical}$) | 32 | 42 | 29 | 23 |
| Total quantity of redundant information ($R_2$) | 3 | 16 | 26 | 9 |
| Total quantity of irrelevant information (IR) | 2 | 7 | 17 | 9 |
| VD[a] | 39 | 65 | 72 | 41 |
| ID[b] | .82 | .65 | .40 | .56 |

*Note*. [a]VD = Total quantity of information = $R_1 + R_2$ + IR. [b]ID = Total quantity of relevant information/total quantity of information

Based on the calculated ID ratio, in total, four display conditions were developed, as shown in Figure 3.1, Figure 3.2, Figure 3.3, and Figure 3.4.

**Figure 3.1**

*Low VD, High ID Display Condition*

**Figure 3.2**

*Low VD, Low ID Display Condition*


**Figure 3.3**

*High VD, High ID Display Condition*

**Figure 3.4**

*High VD, Low ID Display Condition*

## Dependent Variables and Measures

The current study intended to investigate the influence of VD and ID on three dependent variables: SA, workload, and search performance.

### *Situation Awareness*

Participant SA was measured using the SAGAT queries (Endsley, 1995) to assess VD and ID's influence on participant SA. Based on the studies that have compared SA evaluation techniques, different SA measures assess different aspects of SA (Nguyen et al., 2019). For example, SA measurement techniques, such as the Situation Awareness-Subjective Workload Dominance Technique (SA-SWORD; Prinzel et al., 2004) and the SART (Selcon & Taylor, 1990), provide an advantage of ease of implementation, but they pose several limitations. Endsley (1995) states that such rating techniques can be affected by participants performing multiple trials, and the direct self-rating collected at the end of the task can be prone to overgeneralization by the participant. Additionally, Endsley (2019) states that self-report measures of SA tend to deviate from the results of SA measure that more directly quantify SA. The main advantage of SAGAT is that it allows an objective, unbiased index of SA that assesses operator SA across a wide range of elements that are important for SA in a dynamic system. The SAGAT approach freezes or pauses the task at hand and asks the participant to answer SA probes targeted at each level of SA that are administered either verbally or via a computer system (Endsley, 1995). This assessment directly measures SA as it taps into the operator's perceptions rather than infers them from behaviors that many other factors besides SA may influence (Endsley et al., 1998a). According to Endsley and Garland (2000), SAGAT queries should include probes regarding all three levels of SA, including Level 1 (Perception of data), Level 2 (Comprehension of meaning), and Level 3 (Projection of the near future) components. In an aviation setting, specifically for a pilot, these probes could include questions, such as "What is the current altitude of your aircraft?" (Level 1), "Enter the deviation between the current track and desired track." (Level 2), and "Enter the minutes remaining before the aircraft lands" (Level 3).

When the participant was flying the eVTOL aircraft, the simulation was paused, and a white blank paper was immediately placed over the pilot interface, after which the participant was asked to respond to Level 1, Level 2, and Level 3 SAGAT queries without referring to the displays or any other information. The SAGAT queries were administered via Qualtrics on an iPad. According to Endsley (1995), it is suggested to have a first SA freeze at least three to five minutes into the task, and no two SA freezes within the same 60 seconds. The SAGAT has been shown to have high reliability with test-retest Cronbach's alpha of .92 to .98 (Endsley & Bolstad, 1994). Endsley et al. (1999)'s research shows the SAGAT to be a valid measure of SA as it correlates with SME ratings of SA. In addition, research has shown a significant correlation between overall SAGAT scores and search performance (Jones & Endsley, 2004; Salmon et al., 2009). However, as SAGAT queries are context-specific, they must be developed based on the task. Due to the lack of previous research that leveraged the SAGAT freeze probe technique to evaluate SA for an eVTOL pilot interface, Level 1, Level 2, and Level 3 freeze probe queries were developed for the current study. To ensure the validity of the developed SAGAT queries, queries were developed using the SA information requirements for commercial pilots by Endsley et al. (1998). They were evaluated for face and content validity by four SMEs, including one eVTOL test pilot, two helicopter pilots, and one aviation human factors researcher. They reviewed the scenarios, stop points, and queries and provided feedback on the content validity for the scoring parameters (threshold of correct vs. incorrect), and verbiage. Correct responses for all Level 1, Level 2, and Level 3 queries were developed to evaluate the participant's response to the queries, including the threshold for what was considered correct versus incorrect. The participant's SAGAT score was a continuous variable. For each display condition, a sum of correct queries was calculated to get each participant's final SA score (see Appendix F). To mitigate the testing effect, a pool of queries was developed. The participants received a total of four queries for each level for each display condition from a pool of queries. The order in which the participants received a select set of queries throughout their four trials was randomized to mitigate testing and learning effects.

***Workload***

The NASA-Task Load Index (NASA-TLX) was used to measure the participant's mental workload. The NASA-TLX has been frequently used as a self-reported questionnaire on cognitive workload. Six workload dimensions are measured, including mental, physical, temporal, performance, effort, and frustration, on a 20-point scale from low to high. For example, participants are asked, "How mentally demanding was the task?" and "How successful were you in accomplishing what you were asked to do?" (Hart & Staveland, 1988). The final result is a cumulative workload score of responses from all six dimensions. The NASA-TLX has been shown to have high test-retest reliability with a Cronbach's alpha of .83 (Hart & Staveland, 1988). For validity, the NASA-TLX has been shown to correlate with other workload measures and subjective ratings of mental workload, as well as being sensitive enough to detect changes in workload manipulations (Longo, 2018; Rubio et al., 2004). Participants were asked to complete the NASA-TLX after each scenario. The NASA-TLX was administered via Qualtrics on an iPad. The overall workload score, which includes the six dimensions, was analyzed as part of the data analysis. The NASA-TLX questionnaire that was used for the study is provided in Appendix G.

***Search performance***

For a crewed eVTOL operation, one of the crucial search performance parameters is the pilot's ability to locate a specific piece of information from the pilot interface. For an eVTOL pilot to perform a smooth landing, an eVTOL pilot should be able to locate information from the pilot interface, such as other waypoints names, flight plan magenta line, airspace indicators, airports, heliports, and other air traffic markers. To measure the participants' search performance, I asked them to name the final approach fix waypoint during each scenario. A participant's search performance was measured with a stopwatch as the time it took them in seconds to name the final approach fix waypoint after being asked to report it. This was repeated for each scenario. Raw values in seconds for each scenario were recorded in an Excel Spreadsheet, separately for each display condition. A higher value indicated that it took longer for the participant to identify the name of final approach fix waypoint, indicating lower search performance. Conversely, a smaller value meant that the participants were able to identify a relevant piece of information faster, indicating higher search performance. Several studies in the past have utilized search performance as a reliable measure in aviation (Beck et al., 2012; Beck et al., 2010) and in automobile interfaces (Pankok Jr. & Kaber, 2018).

In addition to the three primary dependent variables, a demographic survey was administered to the participants before the beginning of the study to capture the characteristics of the sample. The demographic survey asked the participants to report their age, biological sex, race, and ethnicity to compare the sample with the target population. In addition, participants were asked to report their flight hours, the purpose of using the flight simulator, estimated flight hours in the flight simulator, and experience level using a fixed-based flight simulator to control for individual differences in skills and experience. The list of questions that were included in the demographic survey is provided in Appendix H.

***Manipulation Check and Qualitative Measures***

The manipulation check and qualitative measure were developed to assess the participants' perception on the level of VD and the effectiveness of the ID for each display condition (See Appendix B). Participants rated each display condition's level of clutter, with response options ranging from "Not cluttered at all" to "Extremely cluttered," This question aimed to assess the VD of each display condition. Next, participants rated each display condition's level of effectiveness, with response options ranging from "Not at all effective" to "Extremely effective," which assessed how each display condition through different levels of ID assisted in completing the experimental task. Next, participants described what they liked and disliked about the display conditions and explained why, providing insights into specific features or functionalities that enhanced or hindered their ability to effectively complete the experimental task. This manipulation and feedback approach, combining quantitative ratings and qualitative responses, was utilized to validate the display conditions that were developed with varying levels of VD and ID and provide insights for the results from the inferential statistics.

**Figure 3.5**

*Experimental Testbed Setup*

**Experimental Testbed**

The current study used a custom-built, fixed-based flight simulator to run the experimental task. The desktop-based flight simulator, shown in Figure 3.5, consisted of the following hardware:

- Custom-built central processing unit (CPU) equipped with an Intel i9-10850K CPU, 64 GB of RAM, an NVIDIA RTX 3090 Ti graphics card, and Windows 10 Pro operating system.
- Samsung Odyssey G9 Curved, gaming monitor to display out-of-the-window view for the participants.
- Two RealFlightSim Gear PFD and MFD display panels to simulate the pilot interface of an eVTOL aircraft.

- Logitech X-56 H.O.T.A.S Flight Stick and throttle lever to control and fly the aircraft.

The current study utilized X-Plane simulation software. At the time of this writing, the X-Plane 12 simulator was one of the only off-the-shelf flight simulators that allowed users to fly an eVTOL aircraft. The Beta ALIA-250, an eVTOL aircraft model available in the X-Plane 12 Flight Simulator aircraft catalog, was used for the experimental task. The Beta ALIA-250 is an aircraft currently being developed by BETA Technologies and is one of the first eVTOL aircraft that has completed several hundred hours of flight hours, including a fully electric cross-country flight. Several companies, both charter (BLADE Mobility) and cargo companies (UPS), have signed contracts to purchase the ALIA eVTOL after it is certified by the FAA. The aircraft consists of four pairs of propellors used for vertical take-off and landing and a separate pair of propellors specifically to power the aircraft's forward flight, as shown in Figure 3.6 (actual aircraft) and Figure 3.7 (simulated aircraft).

**Figure 3.6**

*BETA Technologies ALIA-250 eVTOL Aircraft*

*Note.* Prototype demonstrator of the Beta Technologies eVTOL Aircraft. Adapted from Beta Technologies Media Kit, by Brian Jenkins, 2021 (https://photos.app.goo.gl/iGaB8GCffRmkowwD7). In the public domain.

**Figure 3.7**

*Simulated BETA Technologies ALIA-250 eVTOL Aircraft*

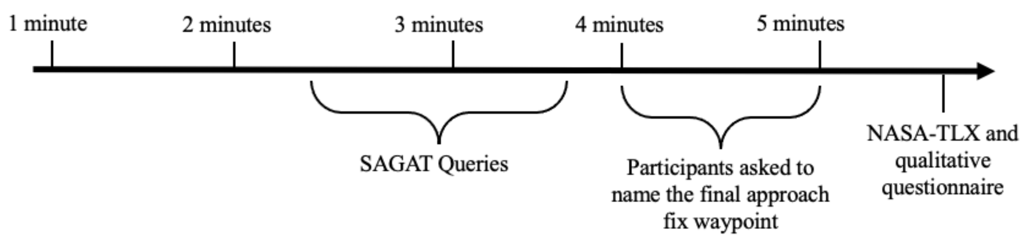
*Note*. This is the eVTOL aircraft simulated in X-Plane 12 flight simulator.

**Experimental Task**

Before beginning the experimental task, the participants were asked to review an eVTOL flight training and task familiarization video. In the video, the participants were provided information about an eVTOL aircraft, shown how to use the flight stick and the throttle lever to fly the eVTOL aircraft, the task they were going to perform, and informed about the SAGAT questions that they needed to answer during the practice flight. After reviewing the eVTOL training video, the participants were given a total of 10 minutes to practice landing the simulated eVTOL aircraft, giving them an opportunity to familiarize themselves with the pilot interface and how the aircraft responded to the control inputs. During this practice session, the participants were administered a set of mock SAGAT queries. After finishing the training flight, the experimental task involved participants completing four scenarios in which they performed visual approaches to Chicago O'Hare International Airport (ORD), Newark Liberty International Airport (EWR), Los Angeles International Airport (LAX), and John F. Kennedy International Airport (JFK) using the testbed described in the previous section. Each of the four scenarios commenced at a distance of 4 nm from each airport. The primary experimental task for the participant flying was to fly and perform a near-vertical landing at the assigned runway at each airport using each of the display conditions. To ensure that no other factors influenced the participant's ability to fly the eVTOL, external environmental conditions, such as wind, were set to ideal conditions. An approximate timeline highlighting the experimental task completed in each scenario is shown in Figure 3.8.

**Figure 3.8**

*Timeline of Experimental Task Completed within each Scenario*




To ensure the experimental task was realistic to the proposed eVTOL operations, several industry SMEs, including one eVTOL test pilot, one air traffic controller, one vertiport development company chief executive officer (CEO), and one eVTOL infrastructure SME, were consulted to ensure the task resembled the currently proposed eVTOL crewed operations. During the development of the experimental task, the SMEs were provided with a detailed description of the tasks the participants would be performing, including the simulator testbed capabilities. One of the SMEs flew one of the scenarios using the simulation testbed setup. Based on the feedback received from the SMEs, the experimental task was modified and was again provided to the SMEs for their feedback. Based on the feedback received, final revisions were made to ensure that the experimental task was realistic to the proposed eVTOL operations.

### *Pilot Testing*

Before the primary data collection, two pilot tests were conducted. The first pilot test was conducted with four SMEs, including two flight instructors and two helicopter pilots, to evaluate the content and face validity of the scenarios. After completing each trial, I queried the SMEs to rate the level of similarity and difficulty for each trial, and whether the experimental task resembled a realistic approach a pilot would perform. The SMEs rated the level of difficulty as moderate (on a scale from extremely easy to very difficult) for each trial and rated the level of similarity as somewhat similar (on a scale from very dissimilar to very similar). Based on the responses provided by the SMEs, two adjustments were made to the scenarios. First, the outside visibility was changed to marginal VFR (MVFR) instead of VFR. This adjustment ensured that the participants used the displays to complete the experimental task. The second adjustment was to begin the approach at a distance of 3 – 4 nautical miles (nm) from the airport, which was previously set at 10 nm, allowing the scenarios to be shortened. After making these adjustments, a second pilot test was then conducted with a sample of five instrument-rated pilots, whose data was used to examine the effect of IV manipulations. All dependent measure data was collected. After this, the means of the dependent variables for each display condition were compared for each independent variable. The analysis of the means revealed no noticeable difference in workload and search performance data. A re-examination of the conditions revealed that a large portion of the display was dedicated to the flight plan page, which did not include any information manipulation other than the size of the information. Based on these comparisons, the four display conditions were adjusted by entirely removing the flight plan page from the MFD panel of the simulation testbed, allowing for a larger portion of the display to be dedicated to use the available information where the variable manipulations occurred. I then used the metrics mentioned in (1) and (2) to recalculate the VD and ID variables and adapted the information on the testbed pilot interface to ensure differences in the quantity and ratio of relevant information across the four display conditions. The revised calculation for the four display conditions is provided in Table 3.3. After revising the display conditions, I resumed pilot testing to re-evaluate the means for the dependent variables. After running four participants, I re-examined the means of the dependent variables for the four display conditions. The results from these four participants revealed noticeable differences across the four display conditions – showing that the revision of the display conditions based on the results of the pilot runs was showing the hypothesized effect of the manipulation on the dependent variables. These four participants were used as my first four official data collection points, and I proceeded with data collection.

***Study Implementation***

To recruit participants, fliers were posted on the FITA flight line, and course instructors were asked to promote the study in aviation classes. In addition, word of mouth was also used to recruit participants. A QR code or hyperlink was provided, which took interested participants to a screening and scheduling page, where they had the opportunity to select the date and time if they met the inclusion criteria of holding an instrument rating for participation in the study. To determine if the participants met the inclusion criteria, prospective participants were required to input their current flight hours and the current pilot license and rating they hold. Then, based on the responses, participants who met the criteria were asked to come to the Florida Institute of Technology's Center for Aeronautics and Innovations (CAI) for the study during the scheduled date and time.

Upon arrival, the participants were given a consent form on an iPad informing them of the study's purpose and associated risks. After which the participants completed a demographics survey, reviewed the training and familiarization video introducing them to eVTOL aircraft and its flying characteristics, and study procedures. Afterward, participants were given the opportunity to practice flying and landing the eVTOL, gaining familiarity with its flight controls and understanding its responsiveness to their inputs. After completing the practice flight, the participants were asked to leave the room, so that the data collector could set up the first display condition of the experimental task. The primary reason for asking the participant to leave the room was to ensure that the participant did not become aware of which display condition they were about to receive. Once the display condition was set, the participants were called into the data collection room to begin the experimental task. Each participant completed four visual approach trials, one in each display condition and one at each of four airports: ORD, EWR, LAX, and JFK. To prevent order effects, the display conditions and the airports were counterbalanced such that each display condition and airport occurred across each of the four trials, and each display condition occurred at each airport. The participants were rotated through these scenario orders using a Latin Square to ensure equal distribution of participants for each of the display conditions (see Appendix I). Randomized level 1, level 2, and level 3 SAGAT queries were administered during each trial. During the experimental task, the simulator was paused, and the participants were asked to complete the SAGAT queries on an iPad. Search performance data was collected by measuring the time it took the participants to name the final approach fix waypoint. After each scenario, participants were asked to leave the data collection room, complete the NASA-TLX questionnaire and fill out the manipulation check questions and the qualitative questionnaire associated with the scenario just completed.

The participants were asked to leave the room after completing the first, second, and third trials. After the participant left the room, the data collector set up the next scenario. At the end of the study, participants were thanked for their participation and given contact information, which they could use to contact the data collector if they had any questions. Once the participant left the data collection room, search performance data for all the display conditions was transferred to an Excel spreadsheet for data analysis. A timeline of the study implementation is shown in Table 3.4.

**Table 3.4**

*Simulation Task Procedure and Timeline*

| Task | Description | Duration |
|---|---|---|
| Introductions | Participants were read an introductory script and asked to sign the consent form | 2 minutes |
| Pre-survey | Participants will complete the demographic survey | 5 minutes |
| Training | Participants reviewed an eVTOL flight training, task familiarization, and introductory eVTOL pilot interface video | 10 minutes |
| eVTOL practice | Participants completed a 10-minute practice flight | 10 minutes |
| Scenario set-up | The proctor set ups the scenario for each display condition (1-minute x 4) | 4 minutes |
| Experimental task | Participants completed a ~6 min experimental scenario for each display condition (6 minutes x 4) | 24 minutes |
| In-task survey | Participants completed an online survey with the SAGAT Queries (2 minutes x 4) | 8 minutes |
| In-task data collection | Participants responded with the name of the final-approach-fix waypoint name for the runway they were flying to. | 1 minute |
| Post-task Survey | Participants completed the qualitative survey and NASA-TLX. (2 minutes x 4) | 8 minutes |
| | ***Total Time*** | ~ 72 minutes |

### *Threats to Internal Validity*

Campbell and Stanley (1963) defined various factors impacting the internal validity of a research design. These factors are defined as threats to internal validity and refer to whether the differences observed in the dependent measures are due to the treatment condition. These threats to internal validity need to be controlled to ensure that the effects on the dependent measures result from the independent variables. A detailed description of all relevant threats applicable to this study are discussed below.

**History.** History threats include cultural or news events that occur during the course of the study that may impact the dependent variable. For example, during data collection, the FAA could publish eVTOL operational procedures or the final version of the eVTOL pilot training requirement, or an eVTOL manufacturer could release details about the pilot interface of their eVTOL aircraft. This could result in participants participating in the study after this event acting more diligently in the experimental task than the previous participants. To control for this effect, changes in the AAM industry were monitored. There was no major event in the AAM industry that was likely to cause a history effect.  However, considering how fast the eVTOL technology is changing, changes in the perceptions of the population regarding AAM and eVTOL aircraft could have occurred over the five months of the studies' data collection period.

**Testing.** A testing effect occurs when the exposure to a pretest alters the participant's search performance of an identical posttest. This effect can cause the participant to become familiar with the task and may perform differently because of the pretest instead of the treatment. In the current study, after a participant completed the first approach at an airport, they might perform better at the second airport due to the exposure to the first visual approach. To control for this effect, the visual approaches the participants performed were set to be similar but at four different airports. In addition to the four different approaches, the primary manipulation of the IVs was counterbalanced with the four airports to mitigate the influence of order effect that may occur.

**Maturation.** A maturation effect occurs when changes occur to the participant over time, such as increased age, decreased motivation, or fatigue. Each participant completed the study in approximately 74 minutes. Previous studies that have conducted experiments of similar timeframe have shown that this duration is not large enough to elicit a significant maturation effect. As a precaution, to ensure that the participants did not exhibit decreased motivation during the four scenarios, the order of the display conditions and the airports the participants received was counterbalanced to wash out any fatigue or motivational effects. Due to the length of the study, the participants were also given the option to take a break between the trials.

**Instrumentation.** An instrumentation effect occurs when changes between measurements occur. This can include differences in the researcher who collects data, biases applied to that data, or interpretations that change as scoring continues. For example, two different researchers may note the search performance of the same individual differently. I was the sole primary data collector for all experimental tasks to control for instrumentation effects. The same dependent measure instruments were used for each participant for all scenarios, the scoring criteria for the SAGAT queries and their scoring criteria was determined before the queries are administered, and an experimenter script and protocol were developed and utilized to ensure each participant received the same the instructions.

**Subject effect.** Two types of subject threats can occur when the participant's perception of the study impacts their responses. These can include an increased search performance from being observed, i.e., the Hawthorne effect, or different performance due to their knowledge of their group assignment, i.e., the John Henry effect. For example, the participant can act differently or perform better if they were aware of being placed in the group that is assigned a less favorable treatment condition or because of the novelty of the treatment. In this study, participants might act more diligently than required because they are observed while performing the experimental task. To mitigate this, the experimenter sat slightly behind the participants to make their presence less conspicuous.

**Experimenter effect.** An experimenter effect can occur when different experimenters administer the treatment differently. For example, a pilot with more experience with flight simulators might be more enthusiastic and involved while administering the treatment than a proctor who is not a pilot. To control for this effect, only I acted as the sole administrator of the study and utilized scripted instructions.

**Location.** A location effect may occur if the study is conducted in many locations. For example, a flight simulation study conducted in a room with other flight simulators or similar equipment may result in different results from the participants. To control the effect of location, the experimental testbed was not moved from the data collection room. Any additional equipment accessories not required for the experimental task were removed from the room for the duration of the study.

**Mortality.** A mortality threat can occur if participants drop out during the study, leading to a more biased sample. Similar simulation studies have been performed in the lab, and dropout rates are typically very low. Nevertheless, any missing data resulting from a participant dropping out was not considered for the primary data analysis.

**Selection.** A selection effect occurs when participants in the control group differ from participants in the treatment group. For example, participants in the group with low levels of VD could show better search performance solely because of individual differences and not because of the treatment condition. For this study, a within-subjects design was used, and participants served as their own control. This controlled for any individual differences. Therefore, this threat to internal validity is not relevant to this study.

**Selection-Maturation.** The selection-maturation threat occurs when participants in different groups mature at different rates. The current study employed a within-subject design where every participant was administered all the treatment conditions. This threat was not relevant to this study as there was only one group.

**Statistical Regression.** The statistical regression effect occurs when participants who score very high or very low on their pretest regress toward the mean on subsequent assessments. If one group has more participants who scored high in the pretest, the data may show that this group had higher flight performance in the posttest than the other group when they did not. This threat was not relevant as a within-subject design was used where participants would serve as their own control.

**Diffusion.** A diffusion effect occurs when participants communicate between groups and learn about the treatment effect. This may result in different behaviors from the participants. For example, a participant who has received the treatment condition might share their experience of flying an eVTOL with one who might have not yet received it. This can alter the participant's behavior as they now know the manipulation and the task. Because a within-subject design is used in the study, participants served as their own control, and this is not a relevant threat to this study. To mitigate the effect of diffusion, participants were asked not to discuss the study with their peers.

## Treatment Verification and Fidelity

Treatment verification and fidelity refer to the extent to which the actual implementation of the study followed the planned study implementations (Shaver, 1983). All the display conditions were manipulated using the specific parameters on the simulation testbed to ensure that the manipulation implementation was identical regardless of the participant. In addition, a checklist was prepared and used, listing every piece of information used for each display condition. All participants eligible to participate in the study received the same demographic survey, study information, and training video. Further, all participants got a chance to complete an eVTOL practice flight. All dependent measures data was collected via Qualtrics and then exported to an Excel spreadsheet. A common script was utilized when introducing the participant to the study to ensure standardization of verbiage outside of the manipulation to ensure treatment fidelity. To ensure ecological validity, inputs from aviation SMEs with experience in eVTOL operations, which included one eVTOL test pilot, two AAM researchers, a consultant, industry reports, current development of eVTOL operations, and ConOps were studied in detail to ensure ecological validity. A detailed description of each is presented in Chapter 3 to ensure that both the independent and dependent variables can be replicated.

**Data Analysis**

Data analysis was performed using both descriptive and inferential statistics. After the data collection, the dependent measure data were exported from Qualtrics and scored accordingly. Demographic data were summarized and presented descriptively by domain, including biological sex, ethnicity, and flight simulator experience for fixed-wing and rotor aircraft. In addition, each measure's descriptive averages and standard deviations are presented to the reader. The participants' SAGAT scores were scored either correct or incorrect within a response correctness range for each query and then added to get the final SA score, resulting in a continuous variable. The NASA-TLX score response was a continuous variable representing the sum of the scores for each question, ranging from 0 to 100, with zero indicating a lower workload and 100 scores indicating the highest workload. Participants' search performance was also a continuous variable. The search performance data was measured as the time taken in seconds by the participant to name the final approach fix waypoint. To determine the effect of the varying levels of VD and ID on the dependent variables, a one-way repeated measure MANOVA was conducted to determine the main effects of, or interaction between, the two independent variables on SA, workload, and search performance data. The data analysis was performed using SPSS 29. Participant's responses to qualitative questions were analyzed separately to evaluate participants' reaction to the display conditions. The results of these analyses are provided in Chapter 4.

# Chapter 4

# Results

## Introduction

This chapter presents the results of the current study. The first section will summarize the descriptive statistics of the sample demographics and dependent and independent variables, including NASA-TLX scores, search performance, and SAGAT scores. The second section will include the results of the inferential statistics, including the preliminary analysis, the one-way MANOVA, and the corresponding univariate results. The preliminary analyses present steps taken to analyze and address invalid or missing data, outliers, and assumptions associated with a MANOVA. In the MANOVA, the results of the multivariate omnibus analysis and univariate analyses for each display condition are provided. The third section presents the results of the participants' responses to the qualitative questionnaire and hypothesis testing results corresponding to the research questions identified in Chapter 1.

## Descriptive Statistics

A total of 26 individuals participated in the study. Before beginning the experimental task, the participants completed a demographic survey. In addition to identifying their demographic information, the demographic survey also asked them to input their experience with flight simulators and their knowledge about advanced air mobility. A summary of the responses to the demographic survey is provided in Table 4.1.

**Table 4.1**

*Summary of the Demographic Variable*

| Demographics | *N* | *M* | *SD* |
|---|---|---|---|
| Biological Sex | | | |
| Male | 20 | | |
| Female | 6 | | |
| Age | 26 | 22.4 | 2.59 |
| Ethnicity | | | |
| Asian | 13 | | |
| African American | 1 | | |
| Mixed | 4 | | |
| White | 8 | | |
| Flight Hours | 26 | 343.14 | 174.50 |
| Academic Level | | | |
| Sophomore | 2 | | |
| Junior | 12 | | |
| Senior | 10 | | |
| Graduate | 2 | | |
| Experience with flight simulator | | | |
| Less than 1 year | 3 | | |
| 1 – 2 years | 13 | | |
| 3 – 4 years | 8 | | |
| More than 4 years | 2 | | |
| Purpose of using flight simulator | | | |
| Instrument flight training | 11 | | |
| Skill development | 9 | | |
| System familiarization | 3 | | |
| Visual flight training | 1 | | |
| All of the above | 2 | | |
| Estimated flight hours in simulator | 26 | 44.07 | 38.89 |
| Familiarity with advanced air mobility | | | |
| Not at familiar | 13 | | |
| Slightly familiar | 11 | | |
| Moderately familiar | 2 | | |

While performing the experimental task using each display condition outlined in Chapter 3, SAGAT query responses and search performance data were collected. After completing the experimental task of landing the eVTOL aircraft for each display condition, the participants were asked to take NASA-TLX and the qualitative questionnaire. The descriptive statistics associated with each dependent variable are presented in the following section.

The SAGAT questionnaire was designed to capture objective SA and was calculated by summing the total number of correct responses to Level 1, Level 2, and Level 3 SAGAT queries. Based on the process outlined by Endsley et al. (1990), a response to a query is deemed correct if it falls within a range that is considered to be operationally close to the correct answer. For example, a correct answer of 70 knots for eVTOL airspeed, a range of 65 – 75 knots, is deemed correct. Both the queries and the ranges of the correct responses were presented to an SME and confirmed to be operationally relevant for an eVTOL aircraft performing an approach at an airport (the experimental task). The range of the correct answers for the developed queries is presented in Appendix F. SAGAT score could range from 0 to 18, with a higher score representing a higher SA. As summarized in Table 4.2 and graphically shown in Figure 4.1, based on the collected responses to the queries, the low VD and high ID display condition had the highest SA ($M$ = 11.69, $SD$ = 1.95), followed by high VD and high ID display condition ($M$ = 10.34, $SD$ = 2.79), low VD and low ID ($M$ = 7.84, $SD$ = 1.69) with the high VD and low ID display condition ($M$ = 7.50, $SD$ = 2.64) having the lowest SA. Generally, SAGAT scores were higher in display conditions that had higher ID.

**Table 4.2**

*SAGAT Scores by Display Conditions*

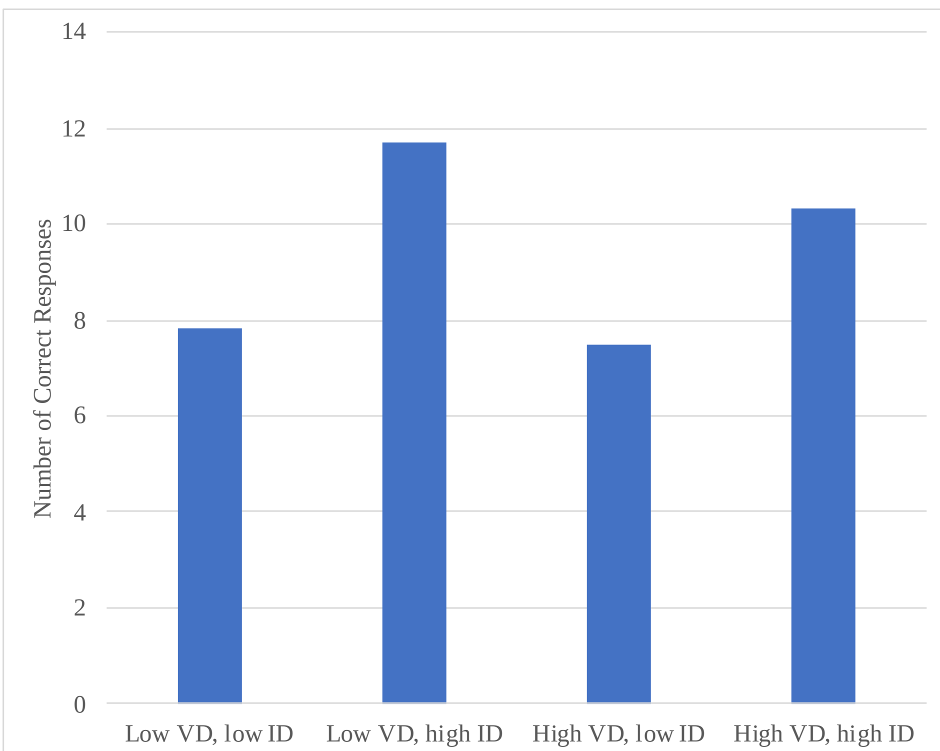
| Display Conditions | *M* | *SD* |
|---|---|---|
| Low VD[a], low ID[b] | 7.84 | 1.69 |
| Low VD, high ID | 11.69 | 1.94 |
| High VD, low ID | 7.50 | 2.64 |
| High VD, high ID | 10.34 | 2.79 |

*Note*. $N = 26$. The Situation Awareness Global Technique (SAGAT) is a measure of real-time situation awareness. The scores could range from 0 to 18, with higher scores representing higher objective situation awareness.

[a] Visual density. [b] Information density.

**Figure 4.1**

*SAGAT Number of Correct Response by Display Conditions*



Workload was measured utilizing the NASA-TLX questionnaire and was administered using Qualtrics on an iPad. The final workload score for each display condition was derived by calculating the average sum of the scores for the six questions for a score range from 0 (lower workload) to 100 (Higher workload).

As summarized in Table 4.3 and shown in Figure 4.2, the low VD and high ID ($M$ = 50.11, $SD$ = 14.80) had a lower workload rating, followed by high VD and high ID ($M$ = 54.11, $SD$ = 19.55), high VD and low ID ($M$ = 58.05, $SD$ = 19.69), with low VD and low ID display condition ($M$ = 59.66, $SD$ = 16.04) reporting the highest workload. Generally, the participants exhibited a lower workload when they were provided with a display condition that consisted of a higher ID.

**Table 4.3**

*Workload Scores by Display Conditions*

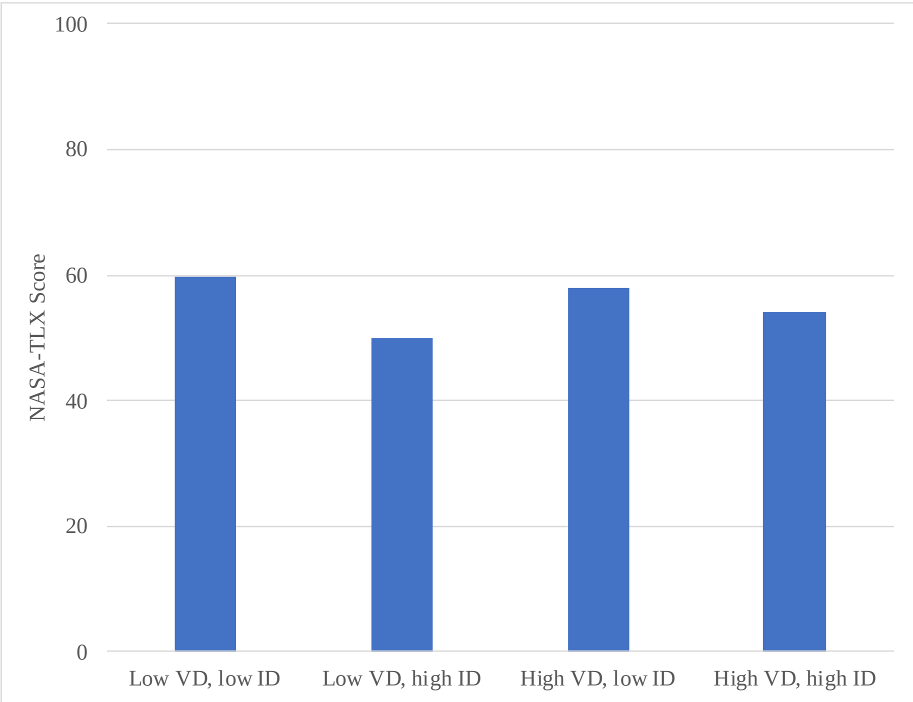
| Display Conditions | *M* | *SD* |
|---|---|---|
| Low VD[a], low ID[b] | 59.66 | 16.04 |
| Low VD, high ID | 50.11 | 14.80 |
| High VD, low ID | 58.05 | 19.39 |
| High VD, high ID | 54.11 | 19.55 |

*Note*. $N$ = 26. The NASA-TLX is a measure of mental workload. Scores could range from 0 to 100, with higher scores representing a higher mental workload.

[a] Visual density. [b] Information density.

**Figure 4.2**

*Workload Rating by Display Conditions*




Search performance was measured by capturing time taken in seconds by participants to name the final approach fix waypoint after being prompted to do so.

As summarized in Table 4.4 and shown in Figure 4.3, search performance was highest with the low VD and high ID display condition (*M* = 4.76, *SD* = 1.79), followed by high VD and high ID condition (*M* = 6.31, *SD* = 1.95), low VD and low ID condition (*M* = 6.56, *SD* = 1.99), with high VD and low ID condition (*M* = 13.4, *SD* = 3.12) showing the lowest search performance. In general, participants spent less time locating a piece of information when they were provided with high ID and/or low VD.

**Table 4.4**

*Search Time by Display Conditions*

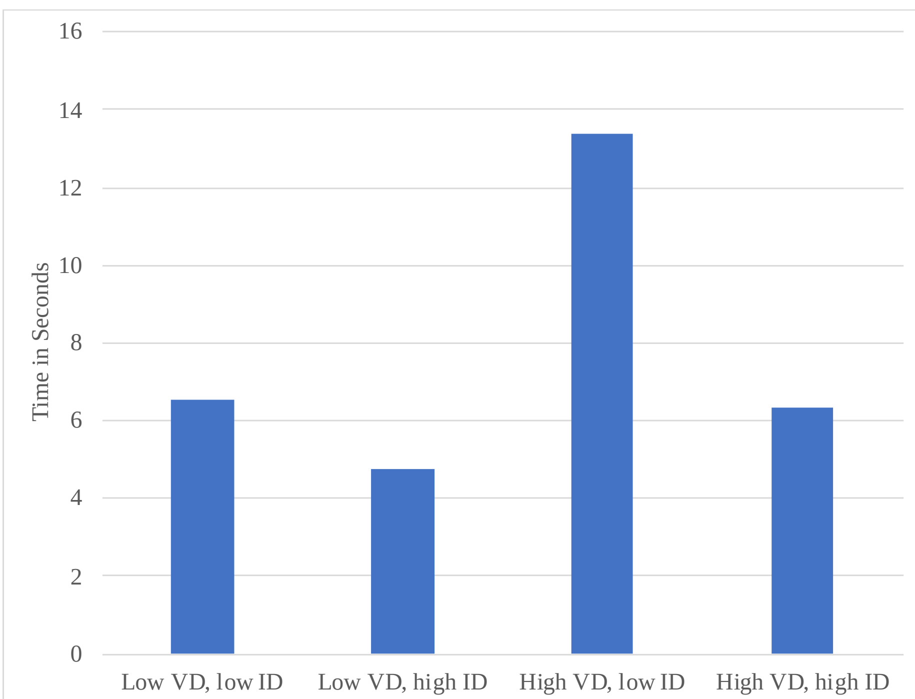
| Display Conditions | *M* | *SD* |
|---|---|---|
| Low VD[a], low ID[b] | 6.56 | 1.99 |
| Low VD, high ID | 4.76 | 1.79 |
| High VD, low ID | 13.4 | 3.12 |
| High VD, high ID | 6.31 | 1.95 |

*Note*. *N* = 26. Search performance was measured by measuring the time in seconds to identify a piece of information from the pilot interface. A higher value represented that it took participants longer to locate the information on the pilot interface.

[a] Visual density. [b] Information density.

**Figure 4.3**

*Search Performance by Display Conditions*



## Inferential Statistics

### *Overview*

The primary purpose of the current study was to explore the influence of VD and ID of an eVTOL aircraft pilot interface on the pilot's SA, workload, and search performance. The research methodology that was best suited to address the research questions associated with the study's purpose was within-subjects repeated measures experimental design. This design was the most appropriate as it allowed the comparison of varying levels of VD and ID on the targeted dependent variables, while also controlling for individual differences. The primary inferential statistical procedure employed for the current study was a repeated measure MANOVA with univariate follow-up analyses.

### *Preliminary Analysis*

**Dataset Modification**. To prepare the data for the analysis, I downloaded the participants' responses to the NASA-TLX questionnaire and the SAGAT queries. The participants' response to search time was tabulated in an Excel spreadsheet. A total of 26 participants took the NASA-TLX questionnaire and responded to the SAGAT queries. After downloading the responses from Qualtrics, the responses were then transferred to an Excel spreadsheet. No other modifications were made to the responses.

**Missing data.** Missing data occurs when participants willingly or unwillingly do not respond to an item on the questionnaire. For search performance, a data point was deemed missing when participants either did not respond when asked to name the final approach fix waypoint or when the participant said they did not understand the question. There were four missing cases for the low VD and low ID, one missing data point for low VD and high ID, three missing data points for high VD and low ID, and one missing data point for high VD and high ID. Upon inspection of the SAGAT data, it was determined that there was one missing data point for the low VD and low ID display condition. According to Cohen et al. (2003), these missing data were replaced with the means of the particular display condition, after confirming that the data points were missing at random and not systematically.

**Outlier Analysis.** In a dataset, outliers are extreme cases that can have an adverse impact on the results of the study. Outliers are cases that lie far away in a dataset compared to the other data points for a particular variable. These outliers can either be contaminated cases or rare cases. In the current study, I performed an outlier analysis using Jackknife distances. A total of eight outliers were detected using this analysis. I further examined each case to determine if they were rare or contaminated cases. Two outliers were detected in the NASA-TLX scores for the low VD, high ID display conditions; upon further inspection, these two data points were less than three standard deviations from the mean, which was determined to be a rare case. These outliers were included in the final data set.

Three outliers were detected in the search time for the low VD, high ID display condition. Upon further inspection, I determined that one participant had taken 13 seconds to name the final approach fix waypoint name. This was more than four standard deviations from the mean of 4.7 seconds. This outlier was deemed as a contaminated data point and removed from the dataset.

Three outliers were detected in the SAGAT score for high VD, high ID display condition. Upon further inspection, it was determined that one participant had scored 18 correct responses, which was more than three standard deviations above the mean of 10 correct responses. This outlier was deemed as a contaminated data point and removed from the data set. Upon further inspection of the remaining two outliers, I determined that they were less than two standard deviations away from the mean. Therefore, these two outliers were deemed as rare cases, and included in the final dataset.

**Multicollinearity.** When a multivariate analysis is conducted, there is a high chance that each variable could be related to one another. When the relationship between the variables is high (for example, $r > .8$), then it could result in a difficult interpretation of the effect of the independent variable on the dependent variable (Cohen et al., 2003). This concept of high correlation between the independent variables is referred to as multicollinearity. To assess the relationship between the variables, I ran bivariate correlations for the dependent variables. The analysis revealed that the correlation ranged from $r = -.10$ to $r = .52$. It was shown that all variables exhibited a correlation of less than .8. Therefore, it was determined that multicollinearity was not an issue.

**Statistical strategy assumptions.** After completing the preliminary data analyses presented above, additional assumptions must be met based on the statistical strategy used for this study. For a MANOVA, the following assumptions must be met: (a) Independence of the dependent variables, (b) linear relationship between the pair of dependent variables, (c) equal variance across the dependent variables, and (d) normal distribution across the dependent variables. Each of the assumptions and their compliance with each assumption are discussed in the sections below.

***Independence of dependent variables.*** The independence assumption is concerned with the observation that each DV is independent of the others. The reader should note that none of the dependent variables for this study, NASA-TLX scores, search performance measured by search time, and SAGAT scores, were dependent on one another. Based on the fact that none of the dependent variables were dependent on one another, this assumption was met.

***Normal distribution.*** This assumption is tested to determine if there is a normal distribution around the mean. To test for normality, a Shapiro-Wilk test for normality was conducted. It was found that, of the three dependent variables, all exhibited a normal distribution except the search performance measure of search time for low VD, high ID and the overall SAGAT score for low VD, high ID display condition. Particularly, for search time data, this was expected due to the low variability discussed in the previous section. Additionally, as suggested by Stevens (2001), "…the sampling distribution of $F$ is only slightly affected, and therefore the critical values when sampling from normal and non-normal distributions will not differ by much" (p. 262). Therefore, the non-normal distribution detected for the two dependent variables did not preclude me from continuing with the primary data analysis.

***Equal variance***. The equal variances assumption is concerned with equal variances across the residuals regardless of the independent variable values. To test this assumption, Levene's test of equality of error variances was conducted. It was found that all DVs satisfied the equality of error variances except for the search time. However, as Stevens (2001, p. 268) notes, "…the *F* statistic is robust against heterogeneous variances when the group sizes are equal." The group sizes were equal for each condition of the dependent measure, given the within-subjects nature of the study. Therefore, non-compliance with the equal variance assumption did not preclude me from continuing with my primary analysis.

***Linearity.*** The linearity assumption is concerned with the type of relationship between the dependent variables. To test this assumption, a bivariate correlation was conducted between each pair of dependent variables. It was discovered that all pairs of the dependent variables exhibited a significant linear relationship except for the NASA-TLX and SAGAT scores, $p > .05$. This was expected as Endsley's (1995a) theory of SA discusses that workload can impact SA. The reader should keep this linear relationship in consideration when interpreting the results of the primary analysis.

***Summary of Preliminary analyses.*** Following the removal of the two outliers during the preliminary analysis, the total sample size included $N = 26$ participants. In terms of missing data, it was determined that the data were missing at random and systematically. A total of five missing data points were identified in the preliminary analyses. Case-wise imputation was used to replace the missing data for the independent variable and the associated measure. No variables were removed due to multi-collinearity, and the independence assumption was met. Equal variance and normality assumptions were violated; however, the primary analyses should not be affected due to the robustness of the *F* test. The linearity assumption violation between the workload and SAGAT score should be noted by the reader when interpreting the results of the primary analysis.

***Primary Analysis***

**MANOVA.** A repeated-measures MANOVA was conducted to examine the effects of VD and ID on SA, workload, and search performance. VD and ID were treated as within-subject factors, and workload (NASA-TLX), search performance (as measured by search time), and SA as the dependent variables. Conducting a MANOVA allowed for an omnibus test to prevent inflation of Type I errors. As shown in Table 4.5, at the multivariate level, there was a significant main effect of VD on the dependent variables, $F(3, 21) = 34.88$, $p < .001$; Wilk's $\Lambda = 0.16$, partial $\eta^2 = .83$. There was also a significant main effect of ID on the dependent variables, $F(3, 21) = 70.85$, $p < .001$; Wilk's $\Lambda = 0.09$, partial $\eta^2 = .91$. Further, the analysis showed significant interactions between VD and ID on the dependent variables, $F(3, 21) = 9.85$, $p < .001$; Wilk's $\Lambda = 0.41$, partial $\eta^2 = .58$. Therefore, univariate follow-up tests were conducted on VD, ID, and the VD*ID interaction.

**Table 4.5**

*Results of Repeated-Measures MANOVA at Multivariate Level*

| Variable | Λ | *F ratio* | *df* | partial $\eta^2$ |
|---|---|---|---|---|
| VD | .167 | 34.882*** | 3, 21 | .833 |
| ID | .090 | 70.850*** | 3, 21 | .910 |
| VD*ID | .415 | 9.851*** | 3, 21 | .585 |

*Note*. $N = 26$

$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

***Univariate Results*.** The results of the follow-up univariate analyses are presented in Table 4.6. The univariate results revealed that VD had a significant effect on search performance, $F(1, 23) = 106.25$, $p < .001$; partial $\eta^2 = .88$. The univariate results also revealed that VD had a significant effect on SAGAT scores, $F(1, 23) = 4.37$, $p = .048$; partial $\eta^2 = .16$ However, there was not a significant effect of VD on workload, $F(1, 23) = 0.08$, $p = .77$, partial $\eta^2 = .004$.

**Table 4.6**

*Results of Repeated-Measures MANOVA at Univariate Level*

| VD | *F ratio* | *df* | *p* | partial $\eta^2$ |
|---|---|---|---|---|
| SAGAT | 4.37 | 1, 23 | .048* | .160 |
| Workload | .08 | 1, 23 | .770 | .004 |
| Search Performance | 106.25 | 1, 23 | .001*** | .822 |
| ID | *F ratio* | *df* | *p* | partial $\eta^2$ |
| SAGAT | 98.29 | 1, 23 | .001*** | .810 |
| Workload | 9.12 | 1, 23 | .006** | .284 |
| Search Performance | 123.71 | 1, 23 | .001*** | .843 |
| VD*ID | *F ratio* | *df* | *p* | partial $\eta^2$ |
| SAGAT | 2.69 | 1, 23 | .114 | .105 |
| Workload | 1.05 | 1, 23 | .316 | .044 |
| Search Performance | 26.68 | 1, 23 | .001*** | .537 |

*Note*. $N = 26$

$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

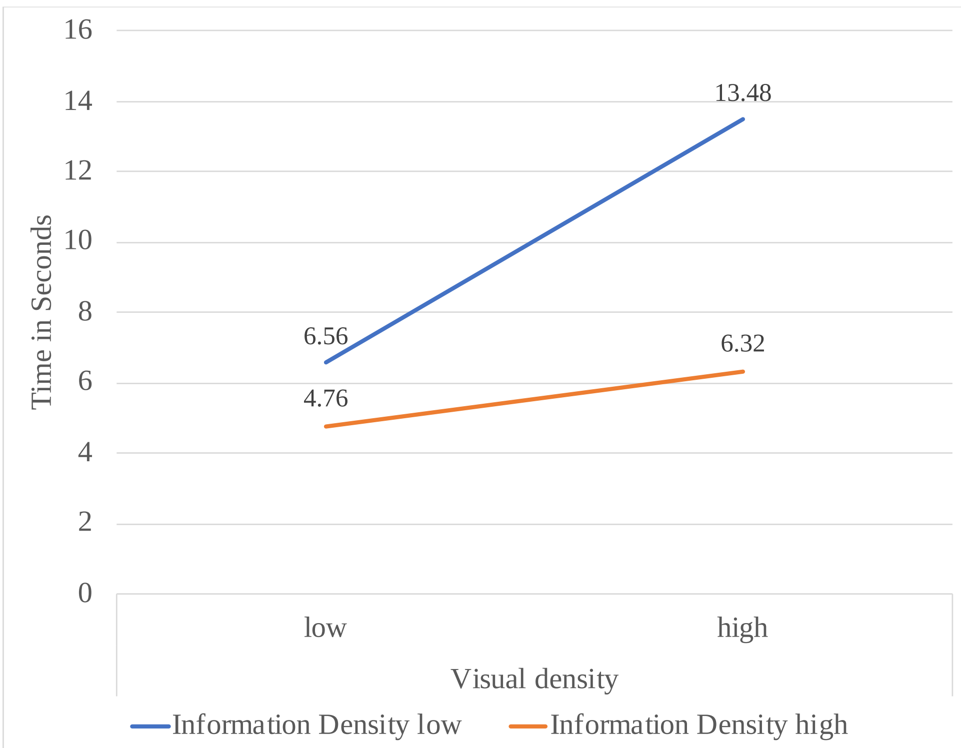
In the context of the current study, pilots using a pilot interface with low levels of VD, i.e., a lesser quantity of information, exhibited higher SA and better search performance compared to when they were using a pilot interface with high VD. The practical significance of the mean differences is discussed in Chapter 5.

In terms of ID, the univariate analysis showed a significant effect of ID on the NASA-TLX scores, $F(1, 23) = 9.13$, $p = .006$; partial $\eta^2 = .28$. In addition, both search performance, $F(1, 23) = 123.71$, $p < .001$; partial $\eta^2 = .84$, and the SAGAT scores, $F(1, 23) = 98.29$, $p < .001$; partial $\eta^2 = .81$ were significantly affected by ID. In the context of the current study, pilots using an eVTOL pilot interface with higher levels of ID, i.e., a higher ratio of relevant information compared to irrelevant and redundant information, experienced lower workload, better search performance, and higher SA when compared to participants using a pilot interface with lower levels of ID.

The follow-up univariate analysis for the interaction between VD and ID revealed one significant variable accounting for the omnibus significance of the interaction. There was a significant interaction between VD and ID on search performance, $F(1, 23) = 26.68$ $p < .001$; partial $\eta^2 = .54$. There was no significant interaction between VD and ID with respect to workload, $F(1, 23) = 1.05$, $p = .32$; partial $\eta^2 = .04$, or SAGAT score, $F(1, 23) = 2.69$, $p = .11$; partial $\eta^2 = .11$. In the context of the current study, these results suggest that there is an interplay between VD and ID on search performance. When VD is low, ID does not have as much of an impact on search performance. However, when there is a high level of VD, the level of ID has a significant impact on search performance, with low ID leading to longer search times than high ID. The reader is reminded that in the current study, VD is defined as the total quantity of information presented on the pilot interface. ID is defined as the ratio of the relevant information on the pilot interface to the total quantity of information. To provide more clarity, the interaction of VD and ID on search performance is shown in Figure 4.4.

**Figure 4.4**

*Interaction of VD and ID on Search Performance*




**Manipulation Check.** To provide a meaningful check of the manipulations for the four display conditions, a series of questions about each display condition were asked to the participants. First, participants were asked to rate the level of clutter on the pilot interface (PFD and MFD combined) on a scale of 1 = “Not at all cluttered" to 5 = "Extremely cluttered". As summarized in Table 4.7 and shown in Figure 4.5, overall, low VD, low ID was rated as the least cluttered, followed by the low VD, high ID display condition, the high VD, high ID display condition, and the high VD, high ID display condition. The participants' responses denote that the low VD display conditions were considerably less cluttered than the high VD display conditions. These findings served as a manipulation check and provided insight into the participant perceptions of the level of visual density for the displays and the associated quantitative findings.

**Table 4.7**

*Participant rating of clutter*

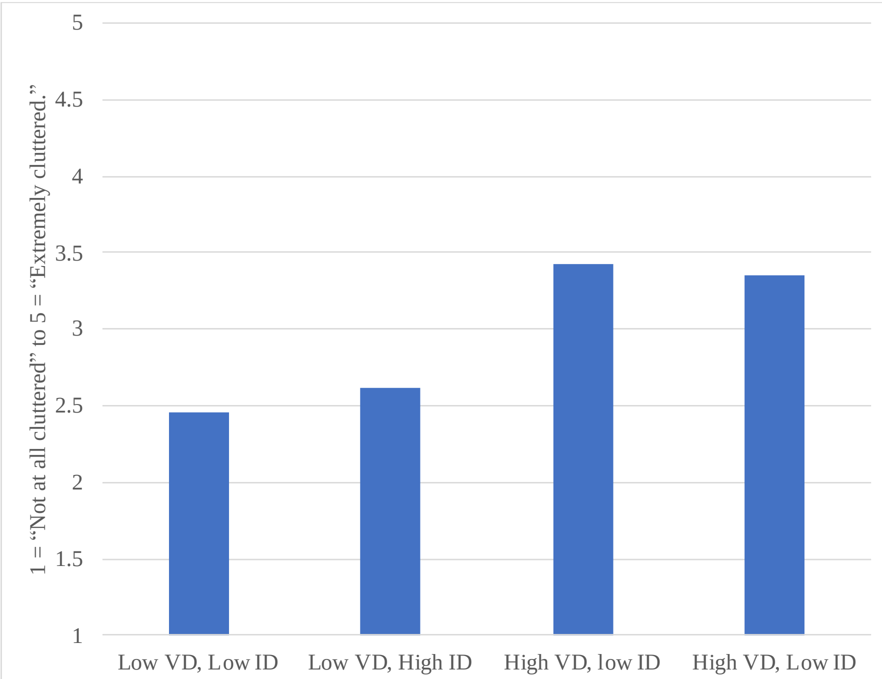
| Display Condition | *M* | *SD* |
|---|---|---|
| Low VD[a], Low ID[b] | 2.46 | 1.03 |
| Low VD, High ID | 2.62 | 1.17 |
| High VD, low ID | 3.42 | .86 |
| High VD, Low ID | 3.35 | 1.02 |

*Note.* The level of clutter was rated on a scale from 1 = "Not at all cluttered" to 5 = "Extremely cluttered."

[a] Visual density. [b] Information density.

**Figure 4.5**

*Rate of Clutter by Display Condition*




Next, the participants were asked to rate the level of effectiveness of the pilot interface in completing the task on a scale from 1 = "Not at all effective" to 5 = "Extremely effective." As summarized in Table 4.8 and shown in Figure 4.6, overall, the participants rated the low VD, high ID display conditions as the most effective, followed by the high VD, high ID display condition, the high VD, low ID display condition, and the low VD, low ID display condition was rated as the least effective.

**Table 4.8**

*Participant rating of the effectiveness of the display conditions*

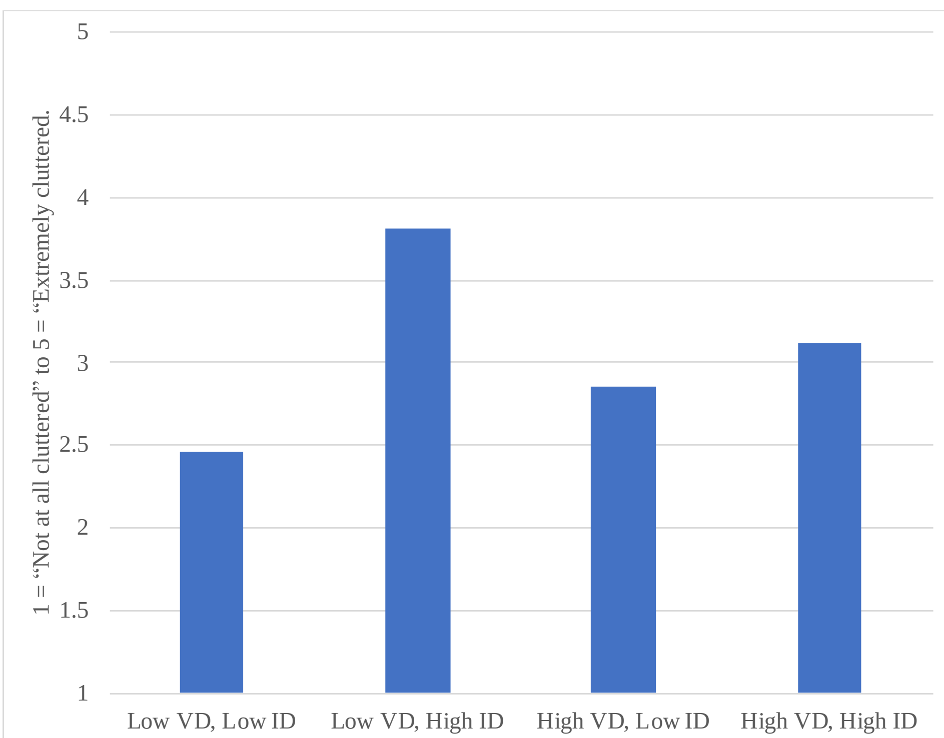
| Display Condition | *M* | *SD* |
|---|---|---|
| Low VD[a], Low ID[b] | 2.46 | .90 |
| Low VD, High ID | 3.81 | .75 |
| High VD, Low ID | 2.85 | .92 |
| High VD, High ID | 3.12 | .82 |

*Note.* The level of effectiveness was rated on a scale from 1 = "Not at all effective" to 5 = "Extremely effective."

[a] Visual density. [b] Information density.

**Figure 4.6**

*Rate of Effectiveness by Display Conditions*



When asked to comment on the level of clutter of the pilot interface, participants reported that the low VD display conditions were less cluttered compared to the high VD display conditions. Notably, for the low VD, low ID display, participants commented that it was "...hard to track important data" and that "information was hard to find." In the low VD, low ID display condition, which had a lower ratio of relevant information compared to the irrelevant and redundant information, participants reported that the "PFD seemed cluttered with unnecessary information" and that "it was harder to find information on the MFD."

The low VD, high ID display condition had a higher ratio of relevant information compared to irrelevant and redundant information. Participants reported that for this display condition, there was "less stuff to distract me on PFD" and "by using the information provided in the interface, I was able to control the aircraft better. For the high VD, low ID display condition, participants reported that the interface "will overwhelm the pilot's workload and divided the attention unevenly." The participants' commented that they had to exercise "extra focus in finding the information instead of the flying the eVTOL aircraft." For the high VD, high ID display condition, the participants reported that the level of clutter did not affect their ability to locate information but made it demanding to "identify information key to the task" and "zone out the unnecessary information and focus only on what's information for me to complete the task." The participants' comments support the inferential results: levels of VD and ID did influence the participant’s ability to find the information presented on the pilot interface to complete the task.

**Qualitative Comments.** Next, the participants were asked to respond to what they liked and disliked about the eVTOL pilot interface. Participants' comments were categorized as either positive comments or negative comments. Then, similar terms were extracted for each display condition from the responses. Table 4.9 and Table 4.10 present the most frequently noted liked and disliked aspects of the display conditions. For the low VD, low ID display conditions, the participants reported that they liked how easy it was for them to see familiar information, which helped them monitor information, and that the display was simple to use. For the low VD, high ID display condition, the participants liked that it was less cluttered with the quality of information that was provided to them to complete the task. For the high VD, low ID display condition, the majority of the participants liked the fact that “it was a G1000 pilot interface”, which they regularly use during their flight training. However, participants also noted that there was too much information. For the high VD, high ID display condition, participants also commented that the eVTOL pilot interface was similar to the G1000 pilot interface, which helped them find information for the task they were performing.

**Table 4.9**

*Qualitative Comment Frequency About What Participants Liked*

| Low VD, Low ID | *f* | Low VD, High ID | *f* | High VD, Low ID | *f* | High VD, High ID | *f* |
|---|---|---|---|---|---|---|---|
| Easy-to-see information | 8 | Ease of use | 4 | Familiar to G1000 | 12 | Similar to G1000 | 8 |
| Clear presentation | 3 | Quality of information | 2 | Quantity of information | 3 | Vertical speed | 7 |
| Easy to monitor | 2 | Less clutter | 2 | | | Easy to find information | 4 |
| Simple display | 2 | RPM information | 2 | | | | |

In terms of dislike for the low VD, low ID display condition, the participants reported there was a “lack of salient information”, making it challenging to adjust to scanning and cross-referencing, hindering their ability to complete tasks. Moreover, participants also reported that the “battery information was inadequate.” For the low VD, high ID, the majority of participants reported that there was “nothing to dislike”. However, participants expressed concern, stating that a first-time eVTOL pilot could find it cluttered. For the high VD, low ID display condition, the participants found it hard to interpret the information presented. Participants identified high VD, high ID display condition to be cluttered.

**Table 4.10**

*Comment Frequency regarding Aspects of the Display the Participants Disliked*

| Low VD, Low ID | *f* | Low VD, High ID | *f* | High VD, Low ID | *f* | High VD, High ID | *F* |
|---|---|---|---|---|---|---|---|
| Lack of salient information | 5 | Nothing to dislike | 5 | Hard to interpret | 9 | Cluttered | 8 |
| Hard to find information | 3 | Couldn’t Zoom in | 3 | Too much information | 7 | | |
| Clutter | 3 | | | | | | |
| Hard to adjust scanning | 2 | Cluttered for first time (eVTOL) pilots | 2 | | | | |

**Results of Hypotheses Testing**

The research questions and hypotheses for the current study are presented in Chapter 1. This section restates the research hypotheses in null form for testing purposes. Each hypothesis is presented along with the corresponding decision to reject or fail to reject.

**Null Hypothesis 1a: There will be no significant effect of varying levels of pilot interface VD on the pilot’s workload**. As shown in Table 4.6, VD had no significant effect on workload. As a result, hypothesis 1a was not rejected, implying that the quantity of information presented on an eVTOL pilot interface has no significant effect on a pilot’s workload.

**Hypothesis 1b: There will be no significant effect of varying levels of pilot interface VD on the pilot's search time.** As shown in Table 4.6, VD had a significant effect on search time. As a result, hypothesis 1b was rejected. This means that the quantity of information presented on an eVTOL pilot interface has a significant effect on a pilot's search time.

**Hypothesis 1c: There will be no significant effect of varying levels of pilot interface VD on pilot SA.** As shown in Table 4.6, VD had a significant impact on SA. As a result, hypothesis 1c was rejected. This implies that the quantity of information presented on an eVTOL pilot interface has a significant effect on a pilot's SA.

**Hypothesis 2a: There will be no significant effect of varying levels of pilot interface ID on the pilot's workload.** As shown in Table 4.6, ID had a significant impact on workload. Therefore, hypothesis 2a was rejected. This shows that the ratio of relevant information to irrelevant and redundant information presented on an eVTOL pilot interface has a significant effect on the pilot's workload.

**Hypothesis 2b: There will be no significant effect of varying levels of pilot interface ID on the pilot's search time.** As shown in Table 4.6, ID had a significant impact on search time. Therefore, hypothesis 2b was rejected. This shows that the ratio of relevant information presented on an eVTOL pilot interface has a significant effect on the pilot's search time.

**Hypothesis 2c: There will be no significant effect of varying levels of pilot interface ID on the pilot's SA.** As shown in Table 4.6, ID had a significant impact on SA. Therefore, hypothesis 2c was rejected implying that the ratio of relevant information presented on an eVTOL pilot interface has a significant effect on the pilot's SA.

**Hypothesis 3a. There will be no significant interaction between VD and ID on workload.** As shown in Table 4.6, there was no significant interaction between VD and ID for workload. Therefore, hypothesis 3a was not rejected. This means that the effect of each IV did not have a significant impact on the effect of the other IV on workload.

**Hypothesis 3b. There will be no significant interaction between VD and ID for search performance.** As shown in Table 4.6, there was a significant interaction between VD and ID for the search performance. Therefore, hypothesis 3b was rejected. This means that when VD was low, ID did not have a large effect; however, when VD was high, ID had a large effect with low ID having significantly higher search times than high ID.

**Hypothesis 3c. There will be no significant interaction between VD and ID for SA.** As shown in Table 4.6, there was no significant interaction between VD and ID for SA. Therefore, hypothesis 3c was failed to be rejected. This means that the effect of each IV did not have a significant impact on the effect of the other IV on SA.

# Chapter 5

# Conclusions, Implications, and Recommendations

## Summary of the Study

The purpose of the current study was to investigate the effect of varying levels of VD and ID of a simulated eVTOL aircraft pilot interface on pilot SA, workload, and search performance during the landing phase of the flight. The independent variables consisted of VD (low vs. high) and ID (low vs. high). The level of VD of the pilot interface was manipulated by changing the total quantity of information presented on the pilot interface. The level of ID was manipulated by changing the ratio of relevant information on the pilot interface to the total quantity of the information. The dependent variables consisted of SA, workload, and search performance. The study utilized a within-subject repeated measures design, which was deemed to be the ideal approach to answer the research question of this study. The order of the VD and ID display conditions were counterbalanced to mitigate order effects. This approach controlled for individual factors, such as experience and previous training.

The target population for the study was all pilots who hold a CPL and operate either scheduled or non-scheduled air taxi and cargo missions. This target population is representative of future eVTOL pilots as they have knowledge of airspace procedures and experience flying at lower altitudes. Further, this population was deemed appropriate for the study based on the preliminary requirement recommended by the FAA for eVTOL pilots to have a CPL and hold an instrument rating. The accessible population included all the student pilots holding a CPL or training to get their CPL enrolled at the Florida Institute of Technology. However, to meet the sample size based on the power analysis, the minimum requirement was reduced to student pilots who have completed their instrument rating. Participants were recruited through word of mouth and online sign-ups using flyers distributed to the flight line and put on Florida Tech's COA bulletin boards. Additionally, the instructors at the COA shared information about the current study in their courses and offered extra credits to any students who participated. Lastly, information regarding the current study also went out on the COA email list to all registered students. Utilizing convenience sampling and the snowball approach, the final sample size obtained was $N = 26$. The demographic breakdown of the sample is presented in Chapter 4 (Table 4.1).

The data collection instruments consisted of (a) objective measurement of SA captured via SAGAT queries, (b) workload captured via the NASA-TLX, and (d) search performance captured via the time taken in seconds to locate the final approach fix waypoint on the pilot interface. The reliability and validity of these measures are presented in Chapter 3.

**Summary of the Findings**

A total of 26 participants completed the study. Before conducting primary data analysis to test the study hypotheses, preliminary data analyses included outlier analysis using Jackknife distance, resulting in a final dataset of $N = 26$, tests for multicollinearity, and MANOVA assumptions testing. A repeated-measures

MANOVA was performed and revealed significant effects of VD and ID on the dependent variables, and an interaction effect between these two variables. A brief summary of the findings and the results of the corresponding hypothesis testing concerning these findings are summarized in Table 5.1.

**Table 5.1**

*Summary of the Results of the Hypothesis Testing*

| Null Hypotheses | | Decision |
|---|---|---|
| 1a | There will be no significant effect of varying levels of pilot interface VD on the pilot’s workload. | Failed to reject |
| 1b | There will be no significant effect of varying levels of pilot interface VD on the pilot’s search performance. | Rejected |
| 1c | There will be no significant effect of varying levels of pilot interface VD on pilot’s SA. | Rejected |
| 2a | There will be no significant effect of varying levels of pilot interface ID on the pilot’s workload. | Rejected |
| 2b | There will be no significant effect of varying levels of pilot interface ID on the pilot’s search performance. | Rejected |
| 2c | There will be no significant effect of varying levels of pilot interface ID on the pilot’s SA. | Rejected |
| 3a | There will be no significant interaction between VD and ID for pilot’s workload. | Failed to reject |
| 3b | There will be no significant interaction between VD and ID for pilot’s search performance. | Rejected |
| 3c | There will be no significant interaction between VD and ID for pilot’s SA. | Failed to reject |

**MANOVA.** A repeated measures MANOVA revealed that at a multivariate level, VD and ID significantly affected the dependent variables. The MANOVA also revealed a significant interaction between VD and ID. A summary of the results of the MANOVA is discussed and presented in Table 4.5. Univariate follow-up analyses revealed that VD significantly affected search performance and

SA but not workload. At the univariate level, ID significantly affected all the dependent variables (see Table 4.6).

**Interactions.** Interactions between VD and ID were revealed for search performance. That is, when VD was low, ID did not have a large effect on search performance, with the low ID condition having very similar search performance as the high ID condition. However, when VD was high, ID had a large effect on search performance, with the low ID condition having a significantly longer search time than the high ID condition. This significant interaction influences how the main effects should be interpreted. Specifically, the interaction indicates that when displays are not visually dense ID does not play a prominent role with respect to search time and may not be a concern.  However, in visually dense displays, it appears to be crucial to provide only the most relevant information to the pilot to maintain optimal search performance. No significant interactions were revealed between VD and ID for workload and SA.

**Manipulation check.** The participants' rating, during the manipulation check measures, of the level of clutter and the level of effectiveness for each display condition suggest that providing a higher ratio of relevant information compared to irrelevant and redundant information removes distraction and supports more effective performance. Participants' responses to the manipulation check questions also served as a validated that the four display conditions were, in fact, perceived to be different in their level of clutter and effectiveness, with the high

VD display conditions being rated as higher in clutter and the high ID conditions being rated as more effective.

**Qualitative comments.** Open-ended responses regarding the participants' opinions about the level of VD and ID revealed subjective opinions regarding how the display conditions impacted them during the experimental task. Particularly, for the low VD and high ID display condition, participants commented that it was simple, intuitive, and easy to see the information, which helped them complete the task. Participants also commented that it was easy for them to see the information on the pilot interface, and "see all the important data in one area" and that it was very clear to look at the pilot interface to get the information. While for the high VD and low ID display condition, participants commented that the pilot interface was too cluttered, due to which they had to divide their attention, they needed to apply extra focus, it was hard to find information, and reported difficulty in monitoring the information on the pilot interface. The higher quantity of information presented on the interface may have contributed to higher levels of workload and made it challenging to focus on the task. The analysis of the qualitative comment for the high VD and high ID display condition followed similar trends to high VD and low ID; participants commented that even though the pilot interface was cluttered, they were able to locate important information, and provided more relevant information about the aircraft which helped them land. These findings imply that that when the participants were provided with a higher quantity of information with a lower ratio of relevant information, it made it difficult to focus on relevant information. Finally, for the low VD and low ID display condition, participants reported that it was “hard to track data” and “information was hard to find.”

**Conclusions and Inferences**

In the following section, the findings from the study are presented and discussed relative to the research questions and terms discussed in Chapter 1. Each section described the results related to the corresponding research question, along with interpretations of those findings in the context of the research setting of the study. Plausible explanations for the findings are also presented.

***Research question 1: What is the effect of pilot-interface VD on pilot SA, workload, and search performance?***

The repeated measures MANOVA revealed a significant effect of VD on search performance and SA, but not a significant effect on workload (see Table 4.6).  Participants experiencing high VD significantly took longer to name the final approach fix waypoint and had significantly lower accuracy on the SA queries. One plausible explanation is that when presented with a visually dense interface, there was an increase in cognitive load as the participant needed to filter and process a higher quantity of information. This added time to the search and could have led to difficulties in effectively locating and interpreting relevant information, impairing search performance. In the context of search performance, a visually dense interface may overwhelm the pilot's visual attention, making it more challenging to identify and locate specific information or cues quickly. This could result in longer search times.

Similarly, regarding SA, a plausible explanation is that a visually dense interface can hinder the pilot's ability to maintain a comprehensive understanding of the current situation. With limited cognitive resources available, the pilot may need help to perceive and comprehend a large quantity of information compared to when the VD is low, leading to a loss of awareness or misinterpretation of the situation.

A plausible explanation for why workload was not significantly impacted is that the experimental task the participants were tasked with completing was performing a visual approach at an airport. Considering the demographic breakdown and the minimum requirements established to participate in the study, the participants were pilots-in-training who had experience performing such approaches during their flight training in an actual fixed-wing aircraft and in-flight simulators. It is possible that due to their proficiency in performing similar tasks to the experimental task in their flight training, the VD of the pilot interface of an eVTOL aircraft did not affect their workload.

***Research Question 2. What is the effect of pilot-interface ID on pilot SA, workload, and search performance?***

The repeated measures MANOVA revealed a significant effect of ID on the dependent variables (see Table 4.6). The results of the current study showed that when ID was low, it led to longer search times, higher workload, and lower SA than when ID was high. A plausible explanation for these results can be grounded in the Broadbent Filter Model of Attention (Broadbent, 1958). That is, when the pilots are fed information from the pilot interface, any information irrelevant to the task at hand must be filtered out, adding workload and time and taking from the cognitive resources necessary to process relevant information.

In the context of the current study, one plausible explanation for the results of the workload based on the Broadbent Filter Model of Attention is that when the participants were presented with a low ID display (where the ratio of relevant information to the redundant and irrelevant information is small), the cognitive load associated with processing information could have increased as the participants needed to exert additional mental effort to filter out information that is not relevant for the task they were performing. This can lead to increased workload.

A plausible explanation for the search performance is that ID directly influenced the participant's ability to locate relevant information on the pilot interface efficiently. When the participants were presented with more relevant information and less irrelevant information, it was easier to locate information on the eVTOL aircraft's pilot interface. However, when the ID is low, the relevant information often gets obscured due to irrelevant and redundant information, which could have led to the participants spending more time looking for relevant information.

A plausible explanation for SA is that in the low ID condition, the participants could have struggled to discern patterns and relationships among the redundant and irrelevant data presented on the pilot interface, especially considering they were flying an eVTOL aircraft. The presence of irrelevant and redundant information could have impeded their ability to form a comprehensive understanding of the current situation – flying an eVTOL aircraft. Conversely, when the participants were provided with more relevant information than redundant and irrelevant information, it could have allowed them to use the relevant information more readily for the task, helping them improve their SA.

***Research Question 3. What is the interaction effect between VD and ID with respect to the pilot's SA, workload, and search performance?***

A repeated measures MANOVA was used to examine the interaction between VD and ID on the dependent variables. The results revealed a significant interaction between VD and ID with respect to the participant's search performance (see Table 4.6). This interaction is the key finding of the current study, as it can influence the interpretation of the main effects of VD and ID on search performance. When VD was low, ID did not have a large effect on search performance, with the low ID condition having very similar search performance as the high ID condition. However, when VD was high, ID had a large effect on search performance, with the low ID condition having a significantly longer search time than the high ID condition. This significant interaction influences how the main effects should be interpreted. Specifically, the interaction indicates that when displays are not visually dense, ID does not significant effect with respect to search performance and it may not be a concern. However, in visually dense displays, the interaction suggest that it is crucial to provide only the most relevant information to the pilot to maintain optimal search performance.

One plausible explanation is that when the display was not visually dense and there was less information presented on the eVTOL pilot interface, in general, it was easy to find the information needed, even if there was a large amount of irrelevant information, likely because there was not as much information to search through. However, when the pilot interfaces are visually dense with information, there is much more information to search through, and therefore, the added increase in irrelevant and redundant information, which can potentially act as a distractor, leading to increased search time.

## Implications

Implications for the current study's results are presented from three aspects: (a) implications relative to theory, (b) implications relative to prior research, and (c) implications for aviation practice.

### *Implications Relative to Theory*

The current study was based on the theoretical foundations of the SEEV model (Wickens et al., 2001) and Broadbent's filter model of attention (1958). In the following section, a brief overview of each theory, a discussion of the implications of the study's findings relative to each theory, and whether the study's findings support or refute the given theory is provided to the reader.

**SEEV Model.** The current study was grounded using the SEEV model proposed by Wickens et al. (2001), which states that the ability to allocate attention is not just limited to the prediction of eye movement. This model identifies four features of visual stimuli that shape attention allocation, incorporating top-down and bottom-up information processing. These four features are (1) Saliency, (2) Effort, (3) Expectancy, and (4) Value. The SEEV model proposes that, among other factors, the effort required to locate information and the relevancy of that information will impact attention allocation processes.

The current study's findings are in line with the SEEV model—notably, the effort required to locate information and the relevancy of information. Based on the model, when the operator is provided with a higher quantity of information, they will require additional time to locate information relevant to the task they are performing, affecting their search performance and ability to maintain SA. The current study identified that when the participants were presented with a visually dense display, it took them longer to identify the name of the final approach fix waypoint on the interface compared to when the VD was low on the eVTOL pilot interface. As the VD on the eVTOL pilot interface increased, the participants had to exert additional effort to locate the waypoint name, resulting in a longer search time, negatively affecting their SA.

Second, the findings of the effect of the levels of ID support the implications of providing relevant information to the operator. Particularly, when the participants were provided with minimal irrelevant information, they were better able to allocate their attention to the information they needed to complete the task. Because of this, the participants experienced a higher SA, lower workload, and better search performance compared to when they were provided with a lower ratio of relevant information to irrelevant and redundant information.

**Broadbent's Filter Model of Attention.** Broadbent (1958), in his Filter Model, proposed that all stimuli are initially processed simultaneously based on their basic physical attributes, such as color, orientation, and saliency. According to Broadbent, during the process of filtering information, when a stimulus is presented, it is initially stored in the sensory store. Subsequently, the information undergoes filtering, wherein a filter acts as the selector of relevant information.

The findings of the current study support Broadbent's Filter Model of Attention. The model states that during the filtration process, only relevant information will be filtered and attended to by the operator. Any irrelevant or redundant information will be filtered out. The ratio of the relevant information to the total quantity of information can determine the cognitive resources the operator will spend filtering relevant information. In the current study, when the participants were provided with a high ID, they showed higher SA, lower workload, and better search performance compared to when they were presented with a low ID pilot interface. These results can be attributed to the fact that when the participants were presented with a higher quantity of irrelevant and redundant information, they had to use excess cognitive resources to filter irrelevant and redundant information from the relevant information, resulting in lower SA, higher workload, and lower search performance.

### *Implications Relative to Prior Research*

The current study was based on prior research findings on VD, ID, and the interaction of VD and ID in aviation and non-aviation domains. The following is a brief overview of the prior research, including those that were both consistent and inconsistent with the current study's results.

**Visual Density.** Current study findings were consistent with basic VD research that has consistently found that higher VD leads to lower search performance, including search time (Bennett et al., 2021; Moacdieh and Sarter, 2017; Van de Weijgert et al., 2019). Further, the results of the current study were consistent with aviation-specific VD research that has found that higher VD in aviation displays led to pilots experiencing lower search performance (Alexander & Wickens, 2005; Backs & Walrath, 1992; Wickens et al., 2005) and lower SA (Beck et al., 2012; Wickens et al., 2004). The literature reviewed that was related to VD did not focus on workload, perhaps because there was not a proposed relationship. As such, the current study findings are consistent with what is currently in the published literature.

**Information Density.** The current study found a significant effect of ID on the SA, workload, and search performance. Participants showed significantly higher SA, lower workload, and better search performance when presented with a high ID eVTOL pilot interface. The current study's findings were consistent with previous research conducted in an aviation context (Alexander et al., 2003; Brahydt & Hansman, 1990; Morphew & Wickens, 1998). Specifically, the results of the current study were consistent with aviation ID research that found that higher ID in aviation displays supported better search performance (Barhydt & Hansman, 1999; Morphew & Wickens, 1998) and lower workload (Alexander et al., 2003). However, the results of the current study, particularly SA, were not consistent with previous studies that found that there was no significant effect of ID on SA (Alexander et al., 2003). One plausible explanation for this inconsistent finding for SA could be attributed to the fact that Alexander et al. evaluated the level of SA by measuring the participant's other air traffic awareness, which was different from how SA was evaluated in the current study.

**Interaction between Visual Density and Information Density.** The current study found a significant interaction between VD and ID on search performance.

The results of the current study were partially consistent with studies that have found an interaction between VD and ID, specifically in regard to search performance, for which it was found that when VD was low, ID did not have a significant effect on performance, however, when VD was high, ID had a large effect, with low ID having significantly lower search performance (Alexander et al., 2008; 2009). This was different from the interaction observed in previous research, which assessed performance through glideslope deviation and found that pilots had the maximum deviation with low VD and high ID (Alexander et al., 2012). However, the workload results of the current study were not consistent with that of the previous research, which showed that when VD was low, ID did have a significant effect on workload (Doyon-Poulin et al., 2014). Lastly, the SA results of the current could not be compared for consistency with previous research, as there were no research studies available that specifically looked into the interaction of VD and ID on the pilot's SA.

***Implications for Aviation Practice***

The implications for the aviation and AAM industry are important to consider, especially due to the emergent nature of the AAM industry and the rapid development of the eVTOL aircraft.

The results of the study show that there is an interaction of VD and ID on a pilot's search performance. Specifically, when displays are visually dense with a higher quantity of irrelevant or redundant information, the pilots will take more time to find critical information, leading to decrease in SA which may impact flight safety, especially during emergencies or time-critical events. With respect to workload, the greater the quantity of information presented to the pilot, the more time and effort they will have to spend to go through all the information to determine which information is critical and help assist with the task. This can lead to a higher workload, which can then further lead to pilots losing sense of their situation and degrading their search performance. With respect to workload, an increased workload associated with displays could take a task that was manageable and make it unmanageable. And reductions in SA can reduce pilot decision making effectiveness. As the eVTOL aircraft pilot interface would be highly customizable, allowing pilots to choose how much and what information to make available on their displays, the findings of the current study can help the industry to understand the effects that this customization may have on pilot search performance and the need to mitigate these effects.

## Generalizability, Limitations, and Delimitations

### *Generalizability*

This section discusses the external validity of the current study, which refers to the extent to which the results of the current study can be applied to other populations and settings beyond that of the current study. In terms of population generalizability, the current study utilized a convenience sampling strategy. As a result, the sample may not be representative of the target population, specifically the current eVTOL pilot requirements of having a commercial license. However, it is likely representative of future pilots, who will be training to fly these vehicles. Efforts are underway to reduce training requirements, making this sample more indicative of the broader future pilot population. Furthermore, it is important to consider the impact of sampling methods. A convenience sample was used, and while students at the Florida Institute of Technology COA come from all over the world, they may not fully represent the entire population of pilots who could eventually become eVTOL pilots. Based on the sample demographics provided in Table 4.1, out of the 26 participants who reported their ethnicity, 13 were Asian, one was African American, four were mixed, and eight were white. Additionally, 20 participants were male, and six were female. Therefore, the study's findings are generalizable to pilots with the demographics described above.

Ecological generalizability refers to the ability for the conditions and the associated results to apply to different settings, conditions, or circumstances. The methods, the eVTOL aircraft, and the experimental task (performing a visual approach at an airport) all impact the ability for the findings of the current study to apply to a real, certified eVTOL aircraft landing at an airport. The reader must take into account the simulated nature of the task, the eVTOL aircraft that was used for the current study, and the airports at which the participants were tasked to land the eVTOL aircraft. First, the simulated task was conducted at four airports (LAX, EWR, ORD, and JFK), which may not be representative of eVTOL flights in other cities or airports. The use of the simulated Lift+Cruise eVTOL aircraft and other features, constraints, and limitations of the current study that are discussed in detail in Chapter 1 can impact the generalizability of the study. However, based on Beta Technologies (2024) and Joby Aviation (2024)'s respective eVTOL flight test campaigns focused on the eVTOL aircraft's landing flight profile, and the results of the study are most applicable to a Lift+Cruise and vectored thrust eVTOL aircraft performing a visual approach at an airport. Another constraint that could limit the ecological generalizability is that the current study was conducted in a simulation, which may not be representative of how pilots will perform in actual eVTOL aircraft or at an airport.

### *Study Limitations and Delimitations*

The current study experienced many limitations and delimitations. For the ease of the reader, the limitations and delimitations from Chapter 1 have been replicated in this section to establish a framework and set the stage for the next section, which presents recommendations for research and practice relative to the study's limitations and delimitations.

**Limitations.** A study's limitations encompass conditions or events beyond the researcher's control that restrict the generalizability of its findings. The limitations of the present study are outlined here, and readers are encouraged to evaluate any conclusions or inferences drawn from the study's results in light of these constraints.

**1. Representativeness of the Sample.** The sample consisted of Florida Tech flight students, who hold an instrument rating. Given that the requirements for future eVTOL pilots do not currently exist, and only provisional pilot training requirements have been made available by the FAA, there may be different training requirements in the near future, yielding additional differences between the proposed sample and the eVTOL pilots, limiting the generalizability of the study.

**2. Representativeness of the Scenarios.** In the current study, the experimental tasks that the participants performed, were based on the review of the FAA AAM Implementation Plan, FAA’s Urban Air Mobility (UAM) ConOps (FAA, 2023d), and recommendations from subject matter experts (SMEs) in aviation with expertise in AAM, aviation planning, air traffic control, and airport operations. As eVTOL aircraft are not certified for commercial operations, the industry does not expect to see for-hire AAM flights for at least the next three to four years. Therefore, modifications in factors relative to the flight, departure, and destination sites can change after the current study is concluded. The experimental task, the flight path, and eVTOL using an active landing runway may not represent future eVTOL flights. This limits the generalizability of the current study. Therefore, future studies that utilize scenarios, such as established AAM flight corridors and vertiports, may yield different results from the current study.

**3. Experience in flying an eVTOL aircraft**. In the current study, participants were tasked to fly an eVTOL aircraft in a simulated environment. As eVTOL aircraft are not yet certified by the FAA for commercial operations or flight training, the sample population will not have any experience flying an eVTOL aircraft, which limits the extent of tasks that I can ask the sample population to perform. Any future study that utilizes certified eVTOL pilots or student pilots training to become eVTOL pilots could yield different results.

**4. Relevant versus irrelevant information**. In the current study, I was limited in identifying relevant and irrelevant information for each of the conditions based on the information that was already displayed on the simulator pilot interface and/or the information that could be customized. Additionally, SMEs were consulted to help determine the relevancy and irrelevancy of the information. A different study that utilizes a different pilot interface or uses different SMEs or a different process to establish relevant and irrelevant information, might yield different results.

**Delimitations.** A study's delimitations are conditions or events that a researcher imposes to make the study more feasible to implement. However, the reader should keep in consideration that these delimitations may further reduce the generalizability of the results. Potential delimitations of the study include:

**1. Sample Strategy.** The current study utilized convenience sampling with the criterion of completion of instrument rating. Using this screening criterion, should allow for control of learning effects and form a homogenous group. Because it is still unclear the flight hours eVTOL pilot would require to fly an eVTOL, a study in the future that uses participants who are in eVTOL flight training or are eVTOL pilots may yield different results.

**2. eVTOL Pilot Interface.** The current study utilized a Garmin G1000 pilot interface that was available with the simulation testbed setup. As stated previously, several eVTOL OEMs have proposed using different Garmin display models, for example, Garmin G3000, for their respective eVTOL aircraft. However, the interface chosen for this study may not accurately represent the pilot interface from a certified eVTOL aircraft but presents information that will be included in an eVTOL aircraft. A study in the future that employs a different pilot interface, for example, an Avidyne, a Honeywell pilot interface, or a Garmin G3000 pilot interface, may yield different results.

**3. Simulated eVTOL Aircraft.** The simulated eVTOL was selected for the current study as it was one of the only available fully functional eVTOL aircraft offered by an off-the-shelf flight simulator. The simulated eVTOL aircraft is an accurate model of an actual eVTOL currently being developed for different applications in the AAM ecosystem. While there are other flight simulators, the X-

Plane 12 testbed was selected due to the availability of an eVTOL aircraft. However, other simulation testbeds with simulated eVTOL aircraft are available for purchase, none of which are representative of an actual eVTOL aircraft. A study employing a different eVTOL aircraft or testbed may yield different results.

**4. Representativeness of the Scenario Challenges.** The experimental tasks developed for the current study scenarios do not span the full range of flight profiles an eVTOL aircraft would be flying. The scenarios for the current study were developed considering the FAA recommendations for initial AAM operations. Studies that use different sets of scenarios that accurately represent the AAM ecosystem, for example, landing at a vertiport in a metropolitan area, might yield different results.

**5. Representativeness of performance measure.** In the current study, I measured the participants' search performance by measuring the time it took them in seconds to name the final approach fix waypoint using the pilot interface. This search performance measure was selected based on past conventional fixed-wing aircraft literature and results from the data analysis pilot run. However, a different study in the future that utilizes a different performance metric or uses a pilot interface from a certified eVTOL aircraft as a measure of search performance could yield different results.

**6. Independent Variable Manipulation.** The current study manipulated the IVs by adding and removing select customizable pieces of information from the pilot interface to develop the four display conditions. Using the available customizability of the panels allowed for a realistic representation of an eVTOL pilot interface. However, a study that utilizes a different method to manipulate VD and ID may yield different results.

**7. SAGAT Queries.** The current study employed queries designed specifically for this mission and simulator context. These queries were developed using the method outlined by Endsley (2000), combined with previously published task analyses and queries, resulting in a total of 20 queries. However, the limited number of queries and the brief task duration may prevent a comprehensive assessment of SA. Additionally, since the queries were self-developed for this study, they have not undergone extensive testing to ensure their validity and reliability. Consequently, the queries may not have provided the most accurate and robust measure of objective SA, and future studies using a different set of queries may yield different results.

**8. Workload Measure.** The current study used the NASA-TLX questionnaire to measure workload. A different study that uses different measures of workload, such as The Bedford Workload Scale (Roscoe, 1984), or physiological measures, such as cardiovascular activity: Heart Rate (HR), Heart Rate Variability (HRV), and Electrocardiography (ECG) may yield different results.

## Recommendations for Research and Practice

### *Recommendations for Research Relative to Study Limitations*

The recommendations for future research relative to study limitations are presented below.

1. The current study utilized pilots who had, at minimum, completed or training to get their instrument rating. Future studies should attempt to obtain a sample that is more representative of the experience and certification that will be required for eVTOL pilots once published by the FAA.
2. The current study utilized a Lift+Cruise eVTOL aircraft for the simulated task. Although representative of some of the eVTOLs in the market, it is not representative of all eVTOL aircraft and potentially the aircraft for future UAM operations. Future studies should utilize other eVTOLs, such as a tilt-wing aircraft.
3. In the current study, the simulated scenario consisted of performing an approach at an airport. Future studies should develop scenarios based on the current industry practice recommendations for eVTOL integration into the National Airspace System (NAS), such as established eVTOL corridors and performing a landing at a vertiport – a dedicated ground infrastructure for eVTOL aircraft.
4. The current study utilized a commercial off-the-shelf flight simulator testbed to host the simulated experimental task. Future studies should

consider using more advanced eVTOL simulation testbeds with more extensive customization and realistic eVTOL flight capabilities.

5. The current study used G1000 to present and manipulate the quantity and ratio of relevant information of the eVTOL flight information to the participants. While G1000 is being considered for future eVTOL applications, future studies should consider other more representative eVTOL pilot interfaces. Also, future studies should consider selecting an interface that allows more information customization that can be representative of an eVTOL aircraft.
6. The current study utilized flight students with very little to zero experience flying an eVTOL aircraft. Future studies should consider participants who have experience flying an eVTOL aircraft.
7. The current study utilized search time to measure the participant's search performance. Future studies should consider using other applicable, eVTOL-specific performance measures, for example, reaction time.
8. Once there are experienced eVTOL pilots, testing to see if the results hold for a population with experience using these interfaces would be interesting.

### *Recommendations for Research Relative to Study Delimitations*

1. The current study utilized convenience and snowball sampling leading to a sample that may not be representative. Future research should utilize an alternative, more robust sampling strategy such as purposive sampling.

2. The current study utilized a Garmin G1000 pilot interface that is available with the simulation testbed setup. Future research should utilize different eVTOL pilot interfaces representative of those proposed by various OEMs, such as Avidyne, Honeywell, or Garmin 3000.
3. The simulated eVTOL was selected for the current study as it was one of the only available fully functional eVTOL aircraft offered by an off-the-shelf flight simulator. However, there are many different types of vehicles and simulation testbeds available. Therefore, future studies should utilize other eVTOLs that may be more representative of UAM.
4. The scenario utilized in the current study did not span the full range of flight profiles or missions an eVTOL aircraft would operate. Future studies should use different sets of scenarios, for example, vertical take-off, hover, transition to forward flight, and vertical landing from a vertiport.
5. The current study measured the time in seconds it took the participants to name the final approach fix waypoint to evaluate search performance. Future studies should use other performance measures, specifically for evaluating flight performance. For example, vertical takeoff performance, or adherence to specific flight procedures such as lateral separation from other air traffic.
6. The current study manipulated VD and ID by adding and removing select pieces of information from the pilot interface of the simulation

testbed. Future studies should consider using pilot interfaces that allow the research to modify the display characteristics of the pilot interface, for example, moving the information from one place to another more robust customizability options to manipulate information presented on the pilot interface of an eVTOL aircraft.

### ***Recommendations for Future Research Relative to Implications***

In this section, a list of recommendations for future research is provided that corresponds with the current study's implications.

1. The current study's findings were consistent with the SEEV model and Broadbent's filter model of attention. Future research should look at replicating the study with the same measures to compare the results.
2. The current study's findings were consistent with the findings from VD research findings that should that higher levels of VD can lead to lower search performance and SA. The results of the current study showed it took participants longer to respond with the name of the final approach fix waypoint when they were presented with a higher quantity of information. The results of the effect of VD on workload were not significant. As discussed in the implication section, this could be due to the similarity of the task of performing an approach at an airport – a flight phase that student pilots use extensively for training. A future study should employ a different

task, such as vertical take-off from a helipad, to examine the effect of VD on the workload.

3. The results of the current study showed that higher ID, characterized by a higher ratio of relevant to irrelevant information, significantly improved SA, reduced workload, and enhanced search performance. The findings of workload and search performance were consistent with previous research. While the findings were inconsistent with the previous research's findings on SA. A future study should use a different set of SAGAT queries or a greater number of queries to examine the impact of ID on the participant's SA.
4. The results of the current study found a significant interaction effect of VD and ID on search performance. A future study should replicate the study to ensure and validate the effect of varying levels of VD and ID on search performance. While the main effects of the effect of VD and ID were presented for completeness, a future study should use a different search performance metric, for example, reaction time to investigate the effect of  interaction of VD and ID on search performance.

***Recommendations for Practice Relative to Implications***

Firstly, considering ID, the current study found a significant effect of ID on the dependent variables. Aviation practitioners, particularly companies that are eVTOL aircraft manufacturers, can tailor cockpit displays to ensure an optimal ratio of relevant information to total information. By minimizing redundant and irrelevant data, pilots can more efficiently process relevant information and maintain situational awareness without compromising their workload and search performance. This may involve employing hierarchical display structures, where relevant information is presented prominently, while information deemed redundant or irrelevant is nested or accessed on demand. Such designs can reduce cognitive load, allowing pilots to allocate attention more effectively and make informed decisions in dynamic flight conditions.

From a regulatory standpoint, federal authorities such as the FAA could establish guidelines for pilot interface design that consider both VD and ID. Regulations should be developed that mandate pilot interfaces that prioritize relevant information while minimizing VD. Compliance with these regulations could be enforced through the certification processes for aircraft and cockpit equipment. By incorporating requirements related to ID and VD into regulatory frameworks, aviation authorities can promote cockpit designs that enhance pilot performance and safety across the industry.

Finally, training programs should incorporate comprehensive modules that educate pilots on the impact of VD and ID on their SA, workload, and search performance. These programs should combine theoretical and practical components to ensure pilots understand how different display configurations can affect their cognitive processes and overall performance. Further pilots can be trained on the impact of visual and ID on their SA, workload, and search performance, such that when they configure their displays, they know to minimize VD and maximize ID.

This study provides seminal findings related to the impact of VD and ID on SA, search performance and workload in the AAM piloting context, in particular, the interesting interaction between ID and VD with respect to search performance. The results of the current study demonstrate that varying levels of VD and ID can significantly affect how well an eVTOL pilot will be able to perceive and use the information presented on the pilot interface. The key finding of this study revealed that there is a significant interaction between VD and ID on a pilot’s search performance. Specifically, when VD is low, low ID does not affect a pilot’s search performance. However, when there is a high VD, ID can have a significant effect on search performance, implying that in a visually dense display, it is essential to minimize the presentation of irrelevant information to the pilot. This interaction emphasizes the importance of carefully balancing VD and ID in the design of eVTOL pilot interfaces to optimize search performance. Given the emerging nature of AAM and eVTOL operations, integrating human factors considerations into the design of pilot interfaces will be essential. This study's findings provide a foundation for eVTOL pilot interface designers to create more effective displays that can improve pilot performance and operational safety.

# References

Ahlstrom, U. (2005). Work domain analysis for air traffic controller weather displays. *Journal of Safety Research, 36*(2), 159–169. https://doi.org/10.1016/j.jsr.2005.03.001

Alcock, C. (2021, April 7). *Avionics supplier claims it can solve eVTOL challenges like power management*. FutureFlight. https://www.futureflight.aero/news-article/2021-04-07/avionics-supplier-claims-it-can-solve-evtol-challenges-power-management

Alexander, A. L., & Wickens, C. D. (2005). Flightpath tracking, change detection and visual scanning in an integrated hazard display. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*(1), 68–72. https://doi.org/10.1177/154193120504900116

Alexander, A. L., Kaber, D. B., Kim, S., Stelzer, E. M., Kaufmann, K., & Prinzel, L. J. (2012). Measurement and modeling of display clutter in advanced flight deck technologies. *The International Journal of Aviation Psychology*, *22*(4), 299–318. https://doi.org/10.1080/10508414.2012.718233

Alexander, A. L., Stelzer, E. M., Kim, S.-H., & Kaber, D. B. (2008). Bottom-up and Top-down Contributors to Pilot Perceptions of Display Clutter in Advanced Flight Deck Technologies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(18), 1180–1184. https://doi.org/10.1177/154193120805201806

Alexander, A. L., Stelzer, E. M., Kim, S., Kaber, D. B., & Prinzel, L. J. (2009). Data and knowledge as predictors of perceptions of display clutter, subjective workload and pilot performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *53*(1), 21–25. https://doi.org/10.1177/154193120905300106

Alexander, A. L., Wickens, C. D., & Hardy, T. J. (2003). Examining the effects of guidance symbology, display size, and Field of view on flight performance and situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*(1), 154–158. https://doi.org/10.1177/154193120304700132

Andre, A. D., & Wickens, C. D. (1989). Proximity Compatibility and Information Display: The Effect of Space and Color on the Analysis of Aircraft Stall Condition (AD-A214 488). U.S. Army Human Engineering Laboratory.

Andritsos, K., Scott, B. I., & Trimarchi, A. (2022). What is in a name: Defining key terms in urban air mobility. *Journal of Intelligent & Robotic Systems*, *105*(4). 81. https://doi.org/10.1007/s10846-022-01694-1

AOPA. (2022, September 9). *Flying the Joby eVTOL simulator* [Video]. YouTube. https://youtu.be/3w6jtKl2ymA

Aragon, C. (2005). Using visualization in cockpit decision support systems. *Proceedings of 2005 IEEE International Conference on Systems, Man and Cybernetics*, *3*, 2396–2401. https://doi.org/10.1109/icsmc.2005.1571507

Archer Aviation. (2022, November 15). *Archer enters into agreement for Garmin to supply key avionic systems for Archer's production aircraft, Midnight™.* Archer Aviation - Investor Relations. https://investors.archer.com/news/news-details/2022/Archer-Enters-Into-Agreement-for-Garmin-to-Supply-Key-Avionic-Systems-for-Archers-Production-Aircraft-Midnight/default.aspx

Arnheim, R. (1970). *Visual Thinking*. Faber.

Bacchini, A., & Cestino, E. (2019). Electric VTOL configurations comparison. *Aerospace*, *6*(3), 26. https://doi.org/10.3390/aerospace6030026

Backs, R. W., & Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, *23*(4), 243–254. https://doi.org/10.1016/0003-6870(92)90152-l

Barhydt, R., & Hansman, R. J. (1999). Experimental studies of intent information on cockpit traffic displays. *Journal of Guidance, Control, and Dynamics*, *22*(4), 520–527. https://doi.org/10.2514/2.4428

Beck, M. R., Trenchard, M., Van Lamsweerde, A., Goldstein, R. R., & Lohrenz, M. (2012). Searching in clutter: Visual attention strategies of expert pilots. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1411–1415. https://doi.org/10.1177/1071181312561400

Beck M.R., Lohrenz M.C., Trafton J.G. (2010). Measuring Search Efficiency in Complex Visual Search Tasks: Global and Local Clutter. *Journal of Experimental Psychology: Applied*, *16*(3), 238-250. https://doi.org/10.1037/a0019633

Bennett, C. R., Bex, P. J., & Merabet, L. B. (2021). Assessing visual search performance using a novel dynamic naturalistic scene. *Journal of Vision*, *21*(1), 5. https://doi.org/10.1167/jov.21.1.5

Broadbent, D (1958). Perception and Communication. London: Pergamon Press.

Camacho, M. J., Steiner, B. A., & Berson, B. L. (1990). Icons vs. alphanumeric in pilot-vehicle interfaces. P*roceedings of the Human Factors Society Annual Meeting*, *34*(1), 11–15. https://doi.org/10.1177/154193129003400104

Campbell, D. T. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally.

Carbonell, J. R. (1966). A queueing model for many-instrument visual sampling. *IEEE Transactions on Human Factors in Electronics*, *HFE-7*, 157–164. https://doi.org/10.1109/THFE.1966.232984

Chang, D., & Nesbitt, K. V. (2006). Identifying commonly-used gestalt principles as a design framework for multi-sensory displays. *Proceedings of 2006 IEEE International Conference on Systems, Man and Cybernetics*. 2452–2457. https://doi.org/10.1109/icsmc.2006.385231

Chauhan, B. B., Osman, M. C., Carmody, K., & Carroll, M. (2022, October). Advanced Air Mobility: Human Factors Considerations for Current eVTOL Pilot Interfaces. In Robert C. Owen (Chair), *National Training Aircraft Symposium* [Symposium]. Embry-Riddle Aeronautical University, Embry-Riddle Aeronautical University. https://commons.erau.edu/ntas/2022/day-3/11

Choe, Y., Lee, Y., Chen, H., & Kim, S. (2021). Look! don't let it weigh you down: The effect of VD on perceived product heaviness and evaluation. *Journal of Business Research*, *126*, 35–47. https://doi.org/10.1016/j.jbusres.2020.12.041

Courtin, C., Mahseredjian, A., Dewald, A., Drela, M., & Hansman, J. R. (2021). A Performance Comparison of eSTOL and eVTOL Aircraft. *American Institute of Aeronautics and Astronautics (AIAA) Forum*. https://doi.org/10.2514/6.2021-3220

Doyon-Poulin, P., Ouellette, B., & Robert, J. (2014). Effects of visual clutter on pilot workload, flight performance and gaze pattern. Proceedings of the *International Conference on Human-Computer Interaction in Aerospace - HCI-Aero '14*, 1–7. https://doi.org/10.1145/2669592.2669656

Doyon-Poulin, P., Robert, J., & Ouellette, B. (2012). Review of visual clutter and its effects on pilot performance: A new look at past research. *In Proceedings of the 2012 IEEE/AIAA 31st Digital Avionics Systems Conference (DASC),* 2D1-1–2D1-11. New York, NY: IEEE. https://doi.org/10.1109/DASC.2012.6382290

Endsley, M. R. (1988). *Situation awareness global assessment technique (SAGAT)*. In *IEEE 1988 National Aerospace and Electronics Conference* (pp. 789–995). IEEE. https://doi.org/10.1109/NAECON.1988.195097

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 65–84. https://doi.org/10.1518/001872095779049499

Endsley, M. R. (2019). A systematic review and meta-analysis of direct objective measures of situation awareness: A comparison of SAGAT and SPAM. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *63*(1), 124–150. https://doi.org/10.1177/0018720819875376

Endsley, M. R., & Bolstad, C. A. (1994). Individual differences in pilot situation awareness. *The International Journal of Aviation Psychology*, *4*(3), 241–264. https://doi.org/10.1207/s15327108ijap0403_3

Endsley, M. R., & Garland, D. J. (2000). Situation Awareness Analysis and Measurement. In *Situation awareness analysis and measurement* (pp. 147–173). Lawrence Erlbaum Associates Publishers.

Endsley, M. R., Farley, T. C., Jones, W. M., Midkiff, A. H., & Hansman, R. J. (1998). *Situation Awareness Information Requirements For Commercial Airline Pilots* (ICAT-98-1). International Center for Air Transportation. http://hdl.handle.net/1721.1/35929

Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998a). A comparative analysis of SAGAT and SART for evaluations of situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(1), 82–86. https://doi.org/10.1177/154193129804200119

Endsley, M., Sollenberger, R., & Stein, E. (1999). The Use of Predictive Displays for Aiding Controller Situation Awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(1), 51–55. https://doi.org/10.1177/154193129904300111

Ewing, G. J., Woodruff, C. J., & Vickers, D. (2006). Effects of local clutter on human target detection. *Spatial Vision*, *19*(1), 37–60. https://doi.org/10.1163/156856806775009232

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Federal Aviation Administration. (2014). *Electronic Flight Displays* (AC 25-11B). https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_25-11B.pdf

Federal Aviation Administration. (2020, April 1). *General information | Federal aviation administration*. FAA. Retrieved July 12, 2023, from https://www.faa.gov/licenses_certificates/airline_certification/135_certification/general_info

Federal Aviation Administration. (2023). *14 CFR Part 141 - Pilot schools* (Docket No. 25910). https://www.ecfr.gov/current/title-14/part-141

Federal Aviation Administration. (2023a). *14 CFR Part 61 - Certification: Pilots, flight instructors, and ground instructors*. https://www.ecfr.gov/current/title-14/part-61

Federal Aviation Administration. (2023b). *Advanced Air Mobility (AAM) Implementation Plan* (Version 1.0). https://www.faa.gov/sites/faa.gov/files/AAM-I28-Implementation-Plan.pdf

Federal Aviation Administration. (2023c). *Integration of powered-lift: Pilot certification and operations* (2023-11497). https://www.federalregister.gov/documents/2023/06/14/2023-11497/integration-of-powered-lift-pilot-certification-and-operations-miscellaneous-amendments-related-to

Federal Aviation Administration. (2023d). *Urban Air Mobility (UAM) Version 2.0 Concept of Operations*. https://www.faa.gov/sites/faa.gov/files/Urban%20Air%20Mobility%20%28UAM%29%20Concept%20of%20Operations%202.0_0.pdf

Garmin. (2021, February 10). *Joby aviation selects G3000 for revolutionary eVTOL aircraft*. Garmin Blog. https://www.garmin.com/en-CA/blog/aviation/joby-aviation-selects-g3000-for-revolutionary-evtol-aircraft/

Government Accountability Office. (2022). *Stakeholders Identified Issues to Address for 'Advanced Air Mobility'* (GAO-22-105020). United States Government Accountability Office. https://www.gao.gov/assets/gao-22-105020.pdf

Goyal, R., Reiche, C., Fernando, C., & Cohen, A. (2021). Advanced air mobility: Demand analysis and market potential of the airport shuttle and air taxi markets. *Sustainability*, *13*(13), 7421. https://doi.org/10.3390/su13137421

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task load index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183. https://doi.org/10.1016/s0166-4115(08)62386-9

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). Effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 210–228. https://psycnet.apa.org/doi/10.1037/0096-1523.25.1.210

Hoh, R. H., Smith, J. C., & Hinton, D. A. (1987). The effects of display and autopilot functions on pilot workload for single pilot instrument flight rule (SPIFR) operations (NASA Contractor Report 4073). Langley Research Center, NASA.

Horrey, W. J., & Wickens, C. D. (2004). Driving and side task performance: The effects of display clutter, separation, and modality. *Human Factors*, *46*, 611–624. https://doi.org/10.1518/hfes.46.4.611.56805

Iani, C., & Wickens, C. D. (2004). Factors affecting task management in aviation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(1), 213–217. https://doi.org/10.1177/154193120404800146

Jones, D. G., & Endsley, M. R. (2004). Use of real-time probes for measuring situation awareness. *The International Journal of Aviation Psychology*, *14*(4), 343–367. https://doi.org/10.1207/s15327108ijap1404_2

Jonsson, J. E., & Ricks, W. R. (1995). *Cognitive Models of Pilot Categorization and Prioritization of Flight-Deck Information* (NASA Technical Paper 3528). National Aeronautics and Space Administration.

Kim, S. H., & Kaber, D. B. (2009, October). Assessing the effects of conformal terrain features in advanced head-up displays on pilot performance. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, *53*(1), 36–40. https://doi.org/10.1177/154193120905300109.

Kinjo, H. (2018). Development Trends and Prospectus for eVTOL: A New Mode of Air Mobility. Mitsui & Co. Global Strategic Studies Institute.

Lazaro, M. J., Kang, Y., Yun, M. H., & Kim, S. (2021). The effects of visual complexity and Decluttering methods on visual search and target detection in cockpit displays. *International Journal of Human–Computer Interaction*, *37*(7), 588-600. https://doi.org/10.1080/10447318.2021.1890491

Lineberger, R., Hussian, A., & Rutgers, V. (2019). *Change is in the Air - The elevated future of mobility: What's next on the horizon?* Deloitte. https://www2.deloitte.com/us/en/insights/focus/future-of-mobility/evtol-elevated-future-of-mobility-summary.html

Littell, J., Putnam, J., & Hardy, R. (2019). The evaluation of composite energy absorbers for use in UAM VTOL vehicle impact attenuation. *Proceedings of the Vertical Flight Society 75th Annual Forum*. https://doi.org/10.4050/f-0075-2019-14553

Lohrenz, M. C., & Hansman, R. J. (2004). Investigating Issues of Display Content vs. Clutter during Air-to-Ground Targeting Missions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(1), 186–190. https://doi.org/10.1177/154193120404800140

Lohrenz, M. C., Trafton, J. G., Beck, M. R., & Gendron, M. L. (2009). A model of clutter for complex, multivariate geospatial displays. *Human Factors*, *51*, 90–101. https://doi.org/10.1177/0018720809333518

Longo, L. (2018). Experienced mental workload, perception of usability, their interaction and impact on task performance. *PLOS ONE*, *13*(8), e0199661. https://doi.org/10.1371/journal.pone.0199661

Mendonca, N., Murphy, J., Patterson, M. D., Alexander, R., Juarex, G., & Harper, C. (2022). Advanced air mobility Vertiport considerations: A list and overview. *AIAA AVIATION 2022 Forum*. https://doi.org/10.2514/6.2022-4073

Moacdieh, N. M., & Sarter, N. (2017). The effects of data density, display organization, and stress on search performance: An eye tracking study of clutter. *IEEE Transactions on Human-Machine Systems*, *47*(6), 886–895. https://doi.org/10.1109/thms.2017.2717899

Moacdieh, N. M., Prinet, J. C., & Sarter, N. B. (2013). Effects of modern primary flight display clutter. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 11–15. https://doi.org/10.1177/1541931213571005

Moacdieh, N., & Sarter, N. (2015). Display Clutter: A Review of Definitions and Measurement Techniques. *Human Factors*, *57*(1), 61–100. https://doi.org/10.1177/0018720814541145

Moore, P., & Fitz, C. (1993). Using gestalt theory to teach document design and graphics. *Technical Communication Quarterly*, *2*(4), 389–410. https://doi.org/10.1080/10572259309364549

Moray, N. (1986). Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman, & J.P. Thomas (Eds.), Handbook of perception and performance, Vol II (pp. 40-1–40-51). New York: Wiley & Sons.

Morphew, M. E., & Wickens, C. D. (1998). Pilot performance and workload using traffic displays to support free flight. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(1), 52–56. https://doi.org/10.1177/154193129804200113

Müller, S., Manzey, D., Schreiter, K., & Luckner, R. (2015). Implementing energy status in head-down cockpit displays. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 926–930. https://doi.org/10.1177/1541931215591270

National Academies of Sciences, Engineering, and Medicine. (2020). *Advancing Aerial Mobility: A National Blueprint*. https://doi.org/10.17226/25646

Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision*, *11*(14), 7. https://doi.org/10.1167/11.14.7

Nguyen, T., Lim, C. P., Nguyen, N. D., Gordon-Brown, L., & Nahavandi, S. (2019). A review of situation awareness assessment approaches in aviation environments. *IEEE Systems Journal*, *13*(3), 3590–3603. https://doi.org/10.1109/jsyst.2019.2918283

Oman, C., Kendra, A., Hayashi, M., Stearns, M., & Burki-Cohen, J. (2001). Vertical navigation displays: Pilot performance and workload during simulated constant angle of descent GPS approaches. *International Journal of Aviation Psychology*, *11*(1), 15–31. https://doi.org/10.1207/S15327108IJAP1101_2

Pankok, C., & Kaber, D. (2018). The effect of navigation display clutter on performance and attention allocation in presentation- and simulator-based driving experiments. *Applied Ergonomics/Applied Ergonomics*, *69*, 136–145. https://doi.org/10.1016/j.apergo.2018.01.008

Pavel, M. D. (2022). Understanding the control characteristics of electric vertical take-off and landing (eVTOL) aircraft for urban air mobility. *Aerospace Science and Technology*, *125*, 107143. https://doi.org/10.1016/j.ast.2021.107143

Peterson, D. A., Uhlarik, J., Raddatz, K. R., & Ward, J. (1999). The role of situational factors and display cues on collision detection and avoidance using a terrain enhanced primary flight display. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(23), 1275–1279. https://doi.org/10.1177/154193129904302303

Prinzel, L. J., Ellis, K. K., Arthur, J. J., Kiggins, D. K., Nicholas, S. N., Lake, R. C., & Ballard, K. (2018). Evaluating synthetic vision displays for enhanced airplane state awareness. *Degraded Environments: Sensing, Processing, and Display 2018*. https://doi.org/10.1117/12.2305191

Rantanen, E., & Talleur, D. (2001). Use of flight recorders for pilot performance measurement and training. *Proceedings of the 2001 International Aviation Training Symposium*. Oklahoma City, OK: FAA Academy.

Rantanen, E., Talleur, D., Taylor, H., Bradshaw, G., Emanuel, T., Lendrum, L., & Hulin, C. (2001). Derivation of pilot performance measures from flight data recorder information. *11th International Symposium on Aviation Psychology*.

Roscoe, A. H. (1984). Assessing pilot workload in flight: Flight test techniques (No. AGARD-CP-373). Neuilly-sur-Seine, France: NATO Advisory Group for Aerospace Research and Development (AGARD)

Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight: A decade of practical use* (90019). Royal Aerospace Establishment. https://apps.dtic.mil/sti/pdfs/ADA227864.pdf

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7*(2), 1–22. https://doi.org/10.1167/7.2.17

Rosner, A., Franke, T., Platten, F., Attig, C. (2019). Eye Movements in Vehicle Control. In: Klein, C., Ettinger, U. (eds) Eye Movement Research. *Studies in Neuroscience, Psychology and Behavioral Economics*. Springer, Cham. https://doi-org.portal.lib.fit.edu/10.1007/978-3-030-20085-5_22

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, *53*(1), 61–86. https://doi.org/10.1111/j.1464-0597.2004.00161.x

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring situation awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics, 39*(3), 490-500. https://doi.org/10.1016/j.ergon.2008.10.010

Sanders, A., & Houtmans, M. (1985). Perceptual processing modes in the functional visual field. *Acta Psychologica*, *58*(3), 251–261. https://doi.org/10.1016/0001-6918(85)90024-1

Scallen, S., Hancock, P., & Duley, J. (1995). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied Ergonomics*, *26*(6), 397–403. https://doi.org/10.1016/0003-6870(95)00054-2

Schons, V. W., & Wickens, C. D. (1993). Visual separation and information access in aircraft display layout (Technical Report ARL-93-7/NASA-A 3 I-93-1). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Schriver, A., & Rantanen, E. (2007). Evaluation of the SEEV model of visual attention allocation in ATC applications. *In 2007 International Symposium on Aviation Psychology* (pp. 625). https://corescholar.libraries.wright.edu/isap_2007/28/

Selcon, S. J., & Taylor, R. M. (1990). *Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design*. In *Situational Awareness in Aerospace Operations* (pp. 5/1–5/8). AGARD, NATO.

Senders, J. (1964). The human operator as a monitor and controller of multidegree of freedom systems. *IEEE Transactions on Human Factors in Electronics, HFE-5*, 2–6. https://doi.org/10.1109/THFE.1964.231647

Smith, J. K., & Caldwell, J. A. (2004). Methodology for Evaluating the Simulator Flight Performance of Pilots (AFRL-HE-BR-TR-2004-0188). United State Air Force Research Laboratory.

Stark, J. M., Comstock, J. R., Prinzel, L. J., Burdette, D. W., & Scerbo, M. W. (2001). A preliminary examination of situation awareness and pilot performance in a synthetic vision environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *45*(2), 40–43. https://doi.org/10.1177/154193120104500208

Van de Weijgert, M., Van der Burg, E., & Donk, M. (2019). Attentional guidance varies with display density. *Vision Research*, *164*, 1–11. https://doi.org/10.1016/j.visres.2019.08.001

Van Geel, A., Borst, C., Van Paassen, M. M., & Mulder, M. (2020). Design and evaluation of vertical situation display reflecting configuration changes. *AIAA SciTech 2020 Forum*. https://doi.org/10.2514/6.2020-1084

Vempati, L., Gawron, V. J., & Winter, S. R. (2023). Advanced air mobility: Systematic review of human factors' scientific publications and policy. *Journal of Air Transportation*, 1–12. https://doi.org/10.2514/1.d0366

Ververs, P. M., & Wickens, C. D. (1998). Head-up displays: Effect of clutter, display intensity, and display location on pilot performance. *The International Journal of Aviation Psychology*, *8*(4), 377–403. https://doi.org/10.1207/s15327108ijap0804_4

Westerbeek, H. G. W., Maes, A., van Deemter, K., Gatt, A., van Gompel, R., & Krahmer, E. J. (2011). Referential scope and visual clutter in navigation tasks. *In Proceedings of the Workshop on the Production of Referring Expressions (PRE-CogSci 2011*) (pp. 1-6). Cogn. Sci. Soc.

Wickens, C. D., & Andre, A. D. (1990). Proximity Compatibility and Information Display: Effects of Color, Space, and Objectness on Information Integration. *Human Factors*, *32*(1), 61–77. https://doi.org/10.1177/001872089003200105

Wickens, C. D., Alexander, A. L., Horrey, W. J., Nunes, A., & Hardy, T. J. (2004). Traffic and flight guidance depiction on a synthetic vision system display: The effects of clutter on performance and visual attention allocation. *PsycEXTRA Dataset*. https://doi.org/10.1037/e577062012-046

Wickens, C. D., Helleberg, J., Goh, J., Xu, X., & Horrey, W. J. (2001). Pilot Task Management: Testing an Attentional Expected Value Model of Visual Scanning (pp. 23): Aviation Research Lab. Institute of Aviation.

Wickens, C. D., Nunes, A., Alexander, A. L., & Steelman, K. (2005). 3D Navigation and Integrated Hazard Display in Advanced Avionics: Workload, Performance, and Situation Awareness Display Clutter Issues in SVS Design (AHFD-05-19/NASA-05-7). NASA Langley Research Center. https://ntrs.nasa.gov/citations/20040077259

Wickens, C., McCarley, J., & Steelman-Allen, K. (2009). NT-SEEV: A model of attention capture and noticing on the flight deck. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *53*(12), 769–773. https://doi.org/10.1177/154193120905301202

Wickens, C.D., & Carswell, C.M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, *37*(3), 473– 494. https://doi.org/10.1518/001872095779049408

Williams, K. W., & Bail, J. D. (2003). *Usability and effectiveness of advanced general aviation cockpit displays for instrument flight procedures* (DOT/FAA/AM-03/17). Federal Aviation Administration.

Woodruff, A., Landay, J., & Stonebraker, M. (1998, May). Constant ID in zoomable interfaces. In *Proceedings of the working conference on Advanced visual interfaces* (pp. 57–65). https://dl.acm.org/doi/pdf/10.1145/948496.948505

Wu, X., Wanyan, X., Zhuang, D., & Liu, S. (2016). Pilot situational awareness modeling for cockpit interface evaluation. *In Engineering Psychology and Cognitive Ergonomics: 13th International Conference, EPCE 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17–22, 2016, Proceedings 13*(pp. 476–484). Springer International Publishing. https://doi.org/10.1007/978-3-319-40030-3_46

Wyrick, A. (2023). From napkin to new heights: Honeywell and vertical aerospace. *Honeywell Aerospace*. https://aerospace.honeywell.com/us/en/about-us/blogs/from-napkin-to-new-heights-honeywell-and-vertical-aerospace

Yang, X., Liu, T., Ge, S., Rountree, E., & Wang, C. (2021). Challenges and key requirements of batteries for electric vertical takeoff and landing aircraft. *Joule*, *5*(7), 1644–1659. https://doi.org/10.1016/j.joule.2021.05.001

Yeh, M., & Wickens, C. D. (2001). Attentional filtering in the design of electronic map displays: A comparison of color coding, intensity coding, and decluttering techniques. *Human Factors*, *43*, 543–562. https://doi.org/10.1518/001872001775870359

Zhang, Y., Sun, Y., & Chen, Y. (2014). A framework for ergonomics design of transport category airplane cockpit. *Procedia Engineering*, *80*, 573–580. https://doi.org/10.1016/j.proeng.2014.09.113

Ziv, G. (2016). Gaze Behavior and Visual Attention: A review of eye tracking studies in Aviation. *The International Journal of Aviation Psychology*, *26*(3–4), 75–104. https://doi.org/10.1080/10508414.2017.1313096

# Appendix A

## IRB



FLORIDA TECH

**RESEARCH INVOLVING HUMAN PARTICIPANTS**
**EXPEDITED/FULL APPLICATION**

This information listed below should be submitted to Florida Tech's IRB if the proposed research has more than minimal risk (none of the exempt conditions apply) or if the research utilizes a special population (children, prisoners, institutionalized individuals, etc.). Please consult the IRB website for detailed information, or contact the IRB Chairperson.
**floridatech.edu/research/compliance--regulations/institutional-review-board**

Submit via email to FIT_IRB@fit.edu.

IRB Contact Information:
Dr. Jignya Patel
IRB Chairperson
FIT_IRB@fit.edu
321-674-7391

**PART 1: GENERAL INFORMATION**

Title of project: Evaluating the Influence of eVTOL Pilot Interface Visual Density and Information Density on Pilot Situation Awareness, Workload and Flight Performance

Date of submission: 11/13/2023

Expected project start date: 12/04/2023    Expected project duration: 1 year

Principal investigator: Bhoomin B Chauhan

Title: Graduate Student

Academic unit: College of Aeronautics

Phone: 307-761-9000    Email: bchauhan2017@my.fit.edu

List all co-investigator(s). Please include name, title, academic unit/affiliation and email.

Bhoomin B Chauhan, Graduate Student, College of Aeronautics, bchauhan2017@my.fit.edu
Dr. Meredith Carroll, Professor, College of Aeronautics, mcarroll@fit.edu

**PART 2: PROJECT SPONSORSHIP INFORMATION**

If any part of this study will be funded by an external funding source (current or planned), you must note the funding source and award/solicitation number below:

This study is not funded.

**PART 3: RESEARCH DESCRIPTION**

1. In lay terms, please describe the GENERAL PURPOSE of the study and how human subjects will be involved. List the SPECIFIC AIMS and RESEARCH QUESTIONS or HYPOTHESES. Avoid the use of jargon when describing the purpose of the study.

Purpose: To examine the effect of varying levels of visual density and information density, on pilot situation awareness (SA), workload, and performance.

Research Question 1. What is the effect of pilot-interface visual density on pilot SA, workload, and performance?
Research Question 2. What is the effect of pilot-interface information density on pilot SA, workload, and performance?
Research Question 3. What is the interaction effect between the levels of visual density and information density with respect to the pilot's SA, workload, and performance?
Research Question 4. What is the participant's reaction to using each of the four display conditions?

Hypothesis 1a: Pilot interfaces with high visual density will lead to lower SA than pilot interfaces with low visual density.
Hypothesis 1b: Pilot interfaces with high visual density will lead to a higher workload than pilot interfaces with low visual density.
Hypothesis 1c: Pilot interfaces with high visual density will lead to lower performance than pilot interfaces with low visual density.
Hypothesis 2a: Pilot interfaces with high information density will lead to higher SA than pilot interfaces with low information density.
Hypothesis 2b: Pilot interfaces with high information density will lead to a lower workload than pilot interfaces with low information density.
Hypothesis 2c: Pilot interfaces with high information density will lead to higher performance than pilot interfaces with low information density.
Hypothesis 3a. There will be an interaction between visual density and information density on SA such that when visual density is low, high levels of information density will lead to increased SA, but when visual density is high, higher levels of information density will lead to lower levels of SA.
Hypothesis 3b. There will be an interaction between visual density and information density on workload such that when visual density is low, high levels of information density will lead to decreased workload, but when visual density is high, higher levels of information density will lead to higher workload.
Hypothesis 3c. There will be an interaction between visual density and information density on performance such that when visual density is low, high levels of information density will lead to increased performance, but when visual density is high, higher levels of information density will lead to lower performance.

2. Outline the INCLUSION CRITERIA for subjects, explaining the rationale for the involvement of any special groups, including children, prisoners, pregnant women or subjects with cognitive impairments. Describe the characteristics of the targeted subjects, including gender, age ranges, ethnic background and health/treatment status. If women or minorities are excluded, provide written justification. Give the number of subjects you anticipate including from each targeted group listed above.

Anticipated sample demographic: Commerically rated pilots or pilots currently training to get commercial pilot license.
Number: Up to 32 pilots
Inclusion criteria: Student pilots who have a commercial pilot license or are in a commercial pilot training program.
Characteristics: Sample will include participants over 18 years of age, of all gender, all ethnic background.

3. Describe sources for potential participants, how subjects will be RECRUITED or the sampling procedures. Attach recruitment advertisement(s), if applicable.

Sampling procedure: Convenience sampling procedure will be used to recruit the sample for the study. To recruit from the accessible population, a flyer will be displayed at the Florida Institute of Technology Flight line and at the College of Aeronautics, as well as emailed out by instructors and posted on social media. Word of mouth will also be used to recruit participants for the study.

4. Describe any COMPENSATION the subjects will receive, including course credit. If monetary compensation is offered, indicate how much the subjects will be paid and describe the terms of payment.

Florida Tech professors, at the College of Aeronautics, may offer class credit for participation in this study. Participants will be entered in a raffle to win a $25 Amazon gift cards.

5. Explain how CONFIDENTIALITY and privacy of participant data (and anonymity if appropriate) will be maintained. If the research study involves collection of images or audio recordings of subjects, explain how the material will be used, who will see the images or hear the recordings and in what setting (refer to the audio/video recording policy).

Participant demographic data will be collected via Qualtrics and will only be accessible to the primary data collector and their advisor. Participant identification data will be stored separately from other information for the purposes of anonymity. Participant data will be kept on a password protected computer and deleted after the conclusion of the study.

FLORIDA TECH

**RESEARCH INVOLVING HUMAN PARTICIPANTS**
**EXPEDITED/FULL APPLICATION**

6. Describe the study design/research/measurement PROCEDURE (e.g., control and experimental groups, etc.). Indicate whether or not the subjects will be randomized for this study. Discuss how you will conduct your study and what measurement instruments you are using. List the specific steps of your research protocol. Explain scientific jargon. Attach a copy of any questionnaires, measurement instruments, interview protocols or a description of topics or an approximate script that will be used. If not available at this time, explain. Please describe your study in enough detail so the IRB can identify what you are doing and why.

This study will utilize a mixed-methods research design. This method is ideal as it will allow me to collect quantitative data regarding the impact of varying levels of visual density and information density on pilot SA, workload, and performance and collect subjective reactions regarding the usability and reaction to using various eVTOL pilot interface display conditions. For the quantitative component, a within-subjects repeated measures design will be used with two independent variables: visual density (low vs. high visual density) and information density (low vs. high information density). This research methodology enables the exploration of quantitative data to identify statistical disparities between the different visual density and information density levels and any potential interactions between the two variables. The study will involve participants with diverse backgrounds and varying levels of experience. A within-subject design will be utilized to account for individual differences, with all participants experiencing each treatment condition. Consequently, there will be a single group of participants in the study. For the qualitative component, a phenomenological-like approach will be utilized by having the participants respond to a series of open-ended questions. The questions will be focused on gathering subjective responses about the usability of the eVTOL pilot interfaces in each condition (See attached).

The proposed study will utilize a custom-built, fixed-based flight simulator (See attached, figure 1). The desktop-based flight simulator, will consist of the following hardware:
- X-Plane 12 flight simulator.
- Custom built central processing unit (CPU) equipped with an Intel i9-10850K CPU, 64 GB of RAM, a NVIDIA RTX 3090 Ti graphics card, and Windows 10 Pro operating system.
- Samsung Odyssey G9 Curved, gaming monitor to display out-of-the-window view for the participants.
- Two RealFlightSim Gear PFD and MFD display panels to simulate the pilot interface of an eVTOL aircraft.
- Logitech X-56 H.O.T.A.S Flight Stick and throttle lever to control and fly the aircraft.

The experimental task will involve participants completing four scenarios in which they will perform visual approaches to Chicago O'Hare International Airport (ORD), Newark Liberty International Airport (EWR), Los Angeles International Airport (LAX), and John F. Kennedy International Airport (JFK) using the simulator testbed (figure 1) with four display conditions of varying levels of visual density and information density. Each of the four scenarios will commence at a distance of three nautical miles from each airport. The order in which the participant performs the approaches to the airport and the display conditions will be counter-balanced to mitigate order and learning effect.

A demographic survey will be used to collect participant's demographic data (See attached, page 2). Participant's SA will be evaluated by calculating the percentage of correct responses to the SAGAT queries (See attached, page 3). Participant's workload will be measured using the NASA-TLX questionnaire (See attached, page 5), participant's reaction to flying an eVTOL aircraft with the varying levels of visual density and information density will be captured using a qualitative questionnaire (See attached, page 6), and participant's flight performance will collected and calculated by the simulator testbed. Data will be collected using Qualtrics and stored in a password-protected laptop.

7. If the study will use deception, describe the nature of the deception, discuss why deception is necessary and fully indicate how participants will be debriefed. Deceptive techniques must be justified by the study's prospective scientific, educational or applied value, and the investigator should explore equally effective alternative procedures that do not use deception. A debrief form/process must be discussed here.

This study will not use deception.

8. Describe all SITES where this research will take place and attach documentation of permission from the appropriate source if the study involves subjects from places other than common public spaces.

This study's data collection will take place in the Center for Aeronautics and Innovation (CAI).

9. Describe any POTENTIAL RISKS (physical, psychological, social, legal or other) and the steps that will be taken to minimize risk. Where appropriate, discuss provisions for ensuring necessary medical or professional intervention in the event of adverse effects to the subjects. Also, where appropriate, describe the provisions for monitoring the data collected to ensure the safety of subjects. Research involving children must carefully assess risks and describe the safeguards in place to minimize these risks.

The risks associated with the study does not exceed of that using a computer or desktop based flight simulator.

10. Discuss the importance of the knowledge that will result from your study and what benefits will accrue to your subjects (if any). Discuss why the risks to subjects are reasonable in relation to the anticipated BENEFITS to subjects.

Significance:
1. Help future air carrier operators identify how the pilot interface of an electric aircraft could influence the pilot's ability to fly the aircraft.

2. Provide insight to aircraft manufacturers to develop more effective interfaces, resulting in safer and more efficient electric aircraft operations.

3. Help inform guidance for this customization or the requirement that will bind how customizable the displays on an eVTOL aircraft can be.

11. CONSENT. Informed consent can be in either written or oral format. If you request waiver of informed consent, documentation of informed consent or of written informed consent, please state your justifications. Attach consent form if applicable. If an oral consent is planned, attach a copy of the text of the statement. If the study will be conducted with minors, provide an assent script. If assent is deemed unnecessary or inappropriate, you must discuss why. The consent form should contain all eight elements listed in Part 4. Researchers are strongly encouraged to use the formal headers found in Part 4, Item 3 to structure the consent document.

Before participating in the study, the participants will be asked to sign a consent form which will inform them about the purpose of the study and risks associated with it. The consent form that will be used for the study is attached with the application.

## PART 4: INSTRUCTIONS FOR DOCUMENTATION OF INFORMED CONSENT

Informed consent is one of the primary ethical requirements underlying human subjects research, reflecting the principle of respect for potential subjects. Informed consent assures that prospective human subjects understand the nature of the research and can decide knowledgeably and voluntarily whether or not to participate.

Informed consent refers to the voluntary choice of an individual to participate in research based on an accurate and complete understanding of, among other things, its purposes, procedures, risks, benefits, alternatives and any other factors that may affect a person's decision to participate.

The basic concepts of the consent process include:

- Full disclosure of the nature of the research and the subject's participation
- Adequate comprehension on the part of the potential subject
- Voluntary choice to participate

Informed consent must be documented by use of a written consent form approved by the IRB and signed by the participant or the participant's legally authorized representative. A copy should be given to the person signing the form. Even though the IRB has approved a consent procedure, it is the investigator's responsibility to ensure that each potential subject understands the information and to take the appropriate steps necessary to gain that comprehension.

Individuals may not be involved as research participants unless a) they understand the information that has been provided and informed consent has been obtained, or b) the IRB has approved a waiver for informed consent.

**REMEMBER: if the participant is under the age of 18, parental consent is required. This includes college students under the age of 18.**

If the research involves the participation of minors (under 18 years of age), read the description of requirements for research involving children. Additional requirements concerning parental consent forms and child assent are discussed.

**Please follow the instructions for documentation carefully.**

1. The consent form should be written in language that the participants can understand. Whenever possible, simple declarative sentences should be used. Ordinary language should explain technical terms.
2. Avoid the use of exculpatory language through which the subject or the representative is made to waive or appear to waive any of his/her legal rights or release the investigator, sponsor or institution or its agents from liability for negligence.
3. Important information that must be included on the consent form:
   a. Purpose of the research.
   b. Procedures to be followed (what will the participants be asked to do? Include physical requirements or experimental procedures if applicable.)
   c. Foreseeable risks or discomforts to the subjects. What are the risks associated with participating and what safeguards are in place? Include the following statement, where appropriate: "In the event of physical injury resulting from the research procedures, no form of compensation is available. Medical treatment may be provided at your expense or at the expense of your health care insurer (i.e., Medicare, Medicaid, private payer) which may or may not provide coverage. If you have questions it is your responsibility to contact your insurer."
   d. Benefits to the subject or others which may reasonably be expected to result.
   e. Alternative procedures or alternatives to participation, if any.
   f. Level of confidentiality of participant records. Is data anonymous? How will data be stored? If audio or visual records are obtained, how will they be maintained? Who will have access to the data?
   g. Primary investigator's contact information. Point of contact for questions or problems related to this study.

h. IRB contact. Also note the study was approved by Florida Institute of Technology's IRB, and list the current IRB chair and his/her contact information for questions about the rights of people who take part in research.

i. Voluntary participation, refusal and withdrawal. Include the following statement: "Participation is voluntary. Refusal to participate will involve no penalty or loss of benefits to which you are otherwise entitled. You may discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled."

j. Signatures, if appropriate. Provide a place for:

   I. Signature of the participant (or his/her legally authorized representative)

   II. Date of signature

**WAIVER OF INFORMED CONSENT**

The IRB may approve a consent procedure that does not include, or which alters, some or all of the elements of informed consent outlined above or waives the requirements to obtain informed consent, provided the IRB finds and documents that the following four conditions have been met:

- The research involves no more than minimal risk to the subjects;
- The waiver or alteration will not adversely affect the rights and welfare of subjects;
- The research could not practicably be carried out without the waiver or alteration; and
- Whenever appropriate, the subjects will be debriefed—provided with additional pertinent information—after they have participated in the study.

FLORIDA TECH

**RESEARCH INVOLVING HUMAN PARTICIPANTS**
**EXPEDITED/FULL APPLICATION**

**PART 5: SIGNATURE ASSURANCE SHEET**

I understand Florida Institute of Technology's policy concerning research involving human participants, and I agree:

1. To accept responsibility for the scientific and ethical conduct of this research study.
2. To obtain prior approval from the Institutional Review Board before amending or altering the research protocol or implementing changes in the approved consent form.
3. To immediately report to the IRB any serious adverse reactions and/or unanticipated effects on subjects which may occur as a result of this study.
4. To complete, on request by the IRB, a Continuation Review form if the study exceeds its estimated duration.

Principal investigator's signature: DocuSigned by: Bhoomin Chauhan  Date: 11/9/2023

Principal investigator's name (print): Bhoomin Chauhan

**ADVISOR ASSURANCES**
***(If primary investigator is a student)***

This is to certify that I have reviewed this research protocol and that I attest to the scientific merit of the study, the necessity for the use of human subjects in the study to the student's academic program and the competency of the student to conduct the project.

Major advisor's signature: DocuSigned by: Meredith Carroll  Date: 11/9/2023

Major advisor's name (print): Meredith Carroll

**ACADEMIC UNIT HEAD**
***(It is the PI's responsibility to obtain this signature.)***

This is to certify that I have reviewed this research protocol and that I attest to the scientific merit of this study and the competency of the investigator(s) to conduct the study.

Academic unit head's signature: DocuSigned by: Debbie Carstens  Date: 11/10/2023

Academic unit head's name (print): Debbie Carstens

**FOR IRB USE ONLY**

IRB approval ____________________ Date ____________

IRB # ____________________

Florida Institute of Technology
Institutional Review Board

**Notice of Expedited Review Status**
**Certificate of Clearance for Human Participants Research**

| | |
|---|---|
| Principal Investigator: | Bhoomin B Chauhan |
| Date: | December 2, 2023 |
| IRB Number: | 23-129 |
| Study Title: | Evaluating the Influence of eVTOL Pilot Interface Visual Density and Information Density on Pilot Situation Awareness, Workload and Flight Performance |

Your research protocol was reviewed and **approved** by the IRB Chairperson. Per federal regulations, 45 CFR 46.110, your study has been determined to involve no more than minimal risk for human subjects. Federal regulations define minimal risk to mean that the probability and magnitude of harm are no more than would be expected in the daily life of a normal, healthy person.

Unless you have requested a waiver of consent, participants must sign a consent form, and the IRB requires you give each participant a copy of the consent form for their records. For online surveys, please advise participants to print out the consent screen for their files.

All data, which may include signed consent form documents, must be retained in a locked file cabinet for a minimum of three years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained on a password-protected computer if electronic information is used. Access to data is limited to authorized individuals listed as key study personnel.

Prompt reporting to the IRB is required in the following conditions:

- Procedural changes increasing the risk to participants or significantly affecting the conduct of the study
- All adverse or unanticipated experiences or events that may have real or potential unfavorable implications for participants
- New information that may adversely affect the safety of participants or the conduct of the study.

This study is approved for one year from the above date. If data collection continues past this date, a Protocol Renewal Form must be submitted.

*High Tech with a Human Touch™*

150 West University Boulevard, Melbourne, FL 32901-6975 • (321) 674-7316

## Appendix B

Manipulation Check and Qualitative Questions

1. Please rate the level of effectiveness of the eVTOL pilot interface (PFD and MFD combined) in helping you find the information needed to complete the task.
   - o Not at all effective
   - o Slightly effective
   - o Somewhat effective
   - o Very effective
   - o Extremely effective

2. What did you like about the displays of the eVTOL aircraft you just utilized and why?______________________________________________________
   ______________________________________________________________________
   ______________________________________________________________________

3. 3. What did you dislike about the displays of the eVTOL aircraft you just utilized and why?_______________________________________________
   ______________________________________________________________________
   ______________________________________________________________________

4. Please rate the level of clutter on the display you just utilized.
   - o Not cluttered at all
   - o Slightly cluttered
   - o Moderately cluttered
   - o Cluttered
   - o Extremely cluttered.

5. How did the level of clutter on the pilot interface (PFD and MFD combined) influence your ability to complete the task?___________________________
   ______________________________________________________________________
   ______________________________________________________________________

**Appendix C**

List of Information Displayed on Simulation Testbed.

| # | PFD | MFD |
|---|---|---|
| 1 | NAV 1 | COM 1 |
| 2 | NAV 2 | COM 2 |
| 3 | COM 1 | Nav 1 |
| 4 | COM 2 | Nav 2 |
| 5 | Next Waypoint | Ground Speed (critical) |
| 6 | Arriving waypoint (critical) | Track (critical) |
| 7 | Distance to Next Waypoint | Direct track (critical) |
| 8 | Bearing | ETE to Waypoint |
| 9 | Airspeed | Primary Propellor RPM 1 |
| 10 | True Airspeed | Primary Propellor RPM 2 |
| 11 | Moving Map w/Waypoints | Primary Propellor RPM 3 |
| 12 | Moving Map w/Airspace | Primary Propellor RPM 4 |
| 13 | Moving Map w/Airports &heliports | Secondary Propellor |
| 14 | Moving Map w/Active flight plan | Battery Pack 1 Temperature |
| 15 | Moving Map w/Traffic | Battery Pack 2 Temperature |
| 16 | Moving Map w/Topo | Battery Pack 3 Temperature |
| 17 | Moving Map w/Terrain | Battery Pack 4 Temperature |
| 18 | Moving Map w/NEXTRAD | Batter Capacity |
| 19 | Wind | Moving Map w/Flight Plan |
| 20 | DME - Nav 1 | Moving Map w/Traffic |
| 21 | Bearing 1 - GPS RWY Distance | Moving Map w/Topo |
| 22 | Bearing 2 - GPS RWY Distance | Moving Map w/Terrain |
| 23 | Autopilot Heading | Moving Map w/Airways Low |
| 24 | Heading compass - Current Heading | Moving Map w/Airways High |
| 25 | GPS Approach - Arrow | Moving Map w/Waypoints |
| 26 | Course Indicator | Moving Map w/Airspace |
| 27 | Altitude | Moving Map w/Airports & heliports |
| 28 | Vertical speed | Flight Plan Narrow with Waypoint |
| 29 | Nearest Airport Information | Flight Plan Narrow w/DTK |
| 30 | Alerts | Flight Plan Narrow w/Distance |
| 31 | Flight Plan Window w/DTK, distance | Flight Plan Narrow w/Altitude |
| 32 | | Flight Plan Wide w/Waypoints |
| 33 | | Flight Plan Wide/DTK |
| 34 | | Flight Plan Wide w/Distance |
| 35 | | Flight Plan Wide w/ETE |
| 36 | | Flight Plan Wide w/ETA |
| 37 | | Flight Plan Wide w/Bearing |

**Appendix D**

Complete List of Relevant, Irrelevant, Redundant, and Removable Information on the PFD.

| **Information on PFD** | Removable (Y/N) | Relevant | Redundant | Irrelevant |
|---|---|---|---|---|
| NAV 1[1] | N | 0 | 1 | 1 |
| NAV 2[1] | N | 0 | 1 | 1 |
| COM 1[1,2] | N | 1 | 1 | 0 |
| COM 2[1] | N | 0 | 1 | 1 |
| Next Waypoint[3] | N | 1 | 1 | 0 |
| Arriving waypoint | N | 1 | 1 | 0 |
| Distance to Next Waypoint[3] | N | 1 | 1 | 0 |
| Bearing[3] | N | 1 | 0 | 0 |
| Airspeed | N | 1 | 0 | 0 |
| True Airspeed | N | 1 | 0 | 0 |
| Moving Map w/Waypoints | Y | 1 | 1 | 0 |
| Moving Map w/Airspace | Y | 0 | 1 | 1 |
| Moving Map w/Airports &heliports | Y | 1 | 1 | 0 |
| Moving Map w/Active flight plan | N | 1 | 1 | 0 |
| Moving Map w/Traffic | Y | 1 | 1 | 0 |
| Moving Map w/Topo | Y | 0 | 1 | 0 |
| Moving Map w/Terrain | Y | 0 | 1 | 0 |
| Moving Map w/NEXTRAD | Y | 0 | 1 | 0 |
| Wind | Y | 0 | 0 | 1 |
| DME – Nav 1 | Y | 0 | 0 | 1 |
| Bearing 1 – GPS RWY Distance | Y | 0 | 0 | 1 |
| Bearing 2 – GPS RWY Distance | Y | 0 | 0 | 1 |
| Autopilot Heading | N | 0 | 0 | 1 |
| Heading compass – Current Heading | N | 0 | 0 | 1 |
| GPS Approach – Arrow | N | 1 | 0 | 0 |
| Course Indicator | N | 1 | 0 | 0 |
| Altitude | N | 1 | 0 | 0 |
| Vertical speed | N | 1 | 0 | 0 |

| **Information on PFD** | Removable (Y/N) | Relevant | Redundant | Irrelevant |
| --- | --- | --- | --- | --- |
| Nearest Airport Information | Y | 0 | 1 | 1 |
| Alerts | Y | 0 | 0 | 1 |
| Flight Plan Window w/DTK, distance | Y | 1 | 1 | 0 |

*Note*: [1]Information on the PFD that was not removable and was therefore covered in black tape during the low VD, high ID display condition. [2]Information that was not removable and was covered with black tap during the high VD, low ID display condition. [3]Information that was not removable and was covered with black tap during the low VD, low ID display condition.

# Appendix E

Complete List of Relevant, Irrelevant, Redundant, and Removable Information on the MFD.

| **Information on MFD** | Removable (Y/N) | Relevant | Redundant | Irrelevant |
|---|---|---|---|---|
| COM 1[2, 3] | N | 1 | 1 | 0 |
| COM 2[1, 2] | N | 0 | 1 | 1 |
| Nav 1[1, 2] | N | 0 | 1 | 1 |
| Nav 2[1, 2] | N | 0 | 1 | 1 |
| Ground Speed | N | 1 | 0 | 0 |
| Track | N | 1 | 0 | 0 |
| Direct track | N | 1 | 0 | 0 |
| ETE to Waypoint[4] | N | 1 | 0 | 0 |
| Primary Propellor RPM 1 | N | 1 | 0 | 0 |
| Primary Propellor RPM 2 | N | 1 | 0 | 0 |
| Primary Propellor RPM 3 | N | 1 | 0 | 0 |
| Primary Propellor RPM 4 | N | 1 | 0 | 0 |
| Secondary Propellor | N | 1 | 0 | 0 |
| Battery Pack 1 Temperature | N | 1 | 0 | 0 |
| Battery Pack 2 Temperature | N | 1 | 0 | 0 |
| Battery Pack 3 Temperature | N | 1 | 0 | 0 |
| Battery Pack 4 Temperature | N | 1 | 0 | 0 |
| Batter Capacity | N | 1 | 0 | 0 |
| Moving Map w/Flight Plan | Y | 1 | 1 | 0 |
| Moving Map w/Traffic | Y | 1 | 1 | 0 |
| Moving Map w/Topo | Y | 1 | 1 | 1 |
| Moving Map w/Terrain | Y | 0 | 1 | 1 |
| Moving Map w/Airways Low | Y | 0 | 1 | 1 |
| Moving Map w/Airways High | Y | 0 | 1 | 1 |
| Moving Map w/Waypoints | Y | 1 | 1 | 0 |
| Moving Map w/Airspace | Y | 0 | 1 | 1 |
| Moving Map w/Airports & heliports | Y | 1 | 1 | 1 |

| **Information on MFD** | Removable (Y/N) | Relevant | Redundant | Irrelevant |
|---|---|---|---|---|
| Flight Plan Narrow with Waypoint | Y | 1 | 1 | 0 |
| Flight Plan Narrow w/DTK | Y | 1 | 1 | 0 |
| Flight Plan Narrow w/Distance | Y | 1 | 1 | 0 |
| Flight Plan Narrow w/Altitude | Y | 1 | 0 | 0 |
| Flight Plan Wide w/Waypoints | Y | 1 | 1 | 0 |
| Flight Plan Wide/DTK | Y | 1 | 1 | 0 |
| Flight Plan Wide w/Distance | Y | 1 | 1 | 0 |
| Flight Plan Wide w/ETE | Y | 1 | 1 | 0 |
| Flight Plan Wide w/ETA | Y | 1 | 0 | 0 |
| Flight Plan Wide w/Bearing | Y | 1 | 1 | 0 |

*Note*. [1]Information that was covered during the low visual density, high ID display condition. [2]Information that was covered during the high visual density, high ID display condition. [3]Information that was covered during the high visual density, low ID display condition. [4]Information that was covered during the low visual density, low ID display condition.

## Appendix F

## SAGAT Queries

| SA Levels | Pilot SA Information Requirement | SAGAT Queries | Dependent Measure | Range of correctness |
|---|---|---|---|---|
| 1 | Aircraft state: Heading | Enter the current heading of your aircraft? | Error in degrees | ± 10 degrees |
| 1 | Aircraft state: Altitude | Enter the current altitude of your aircraft? | Error in feet | ± 100 feet |
| 1 | Aircraft state: Indicated airspeed | Enter the current indicated airspeed of your aircraft? | Error in knots | ± 20 kts. |
| 1 | Aircraft state: Pitch/attitude | Enter the current pitch of your aircraft? | Error in degrees | ± 5 degrees |
| 1 | Aircraft state: Thrust setting | Enter the current RPM of your primary propellors? | Errors in RPM | ± 20 RPM |
| 1 | Aircraft state: Battery power | Enter the current battery level of your aircraft? | Error in KW/Hr. | ± 2 |
| 1 | Aircraft state: Vertical speed | Enter the current vertical speed of your aircraft? | Error in feet/minute | ± 200 FPM |
| 1 | Aircraft state: Ground speed | Enter the current ground speed of your aircraft? | Error in knots | ± 20 kts. |
| 1 | Aircraft state: Distance | Enter estimated distance to the runway? | Error in seconds | 2 - 6 nm |
| 2 | Aircraft State | Enter the number of aircraft on the moving map | Absolute | 2 - 6 |
| 2 | Aircraft State | Enter the direction (left or right) the aircraft must turn to face north | Absolute | Left or right |

| SA Levels | Pilot SA Information Requirement | SAGAT Queries | Dependent Measure | Range of correctness |
|---|---|---|---|---|
| 2 | Flight plan conformance: Altitude deviation | Enter the deviation between the current vertical speed and the minimum required vertical speed. | Error in degrees | ± 200 FPM |
| 2 | Flight plan conformance: Track deviation | Enter the deviation between the current track and desired track. | Error in degrees | ± 10 degrees |
| 2 | Flight plan conformance: Heading deviation | Enter the deviation between the current heading and desired heading. | Error in degrees | ± 10 degrees |
| 2 | Flight plan conformance: Airspeed deviation | Enter the deviation between the current airspeed and desired airspeed. | Error in knots | ± 5 degrees |
| 3 | Flight plan | Enter the estimated time in seconds to the destination | Absolute | ± 20 seconds |
| 3 | Flight plan | Enter the estimated time in seconds to the next waypoint fix | Absolute | ± 10 seconds |
| 3 | Flight plan | Enter the estimated distance in nm to the destination | Absolute | ± 5 nm |
| 3 | Aircraft | Enter the time in seconds to transition for vertical landing. | Absolute | ± 5 seconds |

| SA Levels | Pilot SA Information Requirement | SAGAT Queries | Dependent Measure | Range of correctness |
|---|---|---|---|---|
| 3 | Flight plan | Enter the location of the touchdown on the runway | Absolute | touchdown zone marking, threshold marking, aiming point marking |

# Appendix G

## NASA TLX Questionnaire

### *NASA Task Load Index*

*Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.*

| Name | Task | Date |
|---|---|---|
| | | |

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

## Appendix H

## Demographic Survey

1. Participant ID:___________________________
2. Age: ______________________
3. Biological Sex:
    - o Male
    - o Female
    - o Prefer not to say.
4. Ethnicity:
    - o White
    - o Black or African American
    - o Asian
    - o Native American or Alaska Native
    - o Native Hawaiian or Pacific Islander
    - o Mixed or Multiracial
    - o Other
    - o Prefer not to say
5. Total flight hours in an actual aircraft (open-ended response)
6. Academic level.
    - o Freshmen
    - o Sophomore
    - o Junior
    - o Senior
    - o Graduate
7. What level of experience do you have in using a flight simulator?
    - o Less than 1 year
    - o 1 – 2 years
    - o 3 – 4 years
    - o More than 4 years
8. What is the purpose of your use of a flight simulator?
    - o Skill development
    - o System familiarization
    - o Instrument flight training
    - o Visual flight training
    - o Other: __________________
9. Estimated flight hours accumulated in a fixed-wing crewed aircraft in a flight simulator. (open-ended response)
10. Estimated flight hours accumulated in a helicopter/rotorcraft in a flight simulator. (open-ended response)
11. How familiar are you with advanced air mobility and eVTOL aircraft?
    - o Not at all familiar
    - o Slightly familiar
    - o Moderately familiar
    - o Very familiar
    - o Extremely familiar

# Appendix I

## Counterbalance Order

| Participant | Order | Trial 1 | Trial 2 | Trial 3 | Trial 4 |
|---|---|---|---|---|---|
| 1 | 1 | LL, EWR | LH, JFK | HL, LAX | HH, ORD |
| 2 | 2 | HH, JFK | LL, LAX | LH, ORD | HL, EWR |
| 3 | 3 | HL, LAX | HH, ORD | LL, EWR | LH, JFK |
| 4 | 4 | LH, ORD | HL, EWR | HH, JFK | LL, LAX |
| 5 | 5 | LL, LAX | LH, EWR | HL, JFK | HH, ORD |
| 6 | 6 | LH, EWR | HL, JFK | HH, ORD | LL, LAX |
| 7 | 7 | HL, JFK | HH, ORD | LL, LAX | LH, EWR |
| 8 | 8 | HH, ORD | LH, LAX | HL, JFK | LL, EWR |
| 9 | 1 | LL, EWR | LH, JFK | HL, LAX | HH, ORD |
| 10 | 2 | HH, JFK | LL, LAX | LH, ORD | HL, EWR |
| 11 | 3 | HL, LAX | HH, ORD | LL, EWR | LH, JFK |
| 12 | 4 | LH, ORD | HL, EWR | HH, JFK | LL, LAX |
| 13 | 5 | LL, LAX | LH, EWR | HL, JFK | HH, ORD |
| 14 | 6 | LH, EWR | HL, JFK | HH, ORD | LL, LAX |
| 15 | 7 | HL, JFK | HH, ORD | LL, LAX | LH, EWR |
| 16 | 8 | HH, ORD | LH, LAX | HL, JFK | LL, EWR |
| 17 | 1 | LL, EWR | LH, JFK | HL, LAX | HH, ORD |
| 18 | 2 | HH, JFK | LL, LAX | LH, ORD | HL, EWR |
| 19 | 3 | HL, LAX | HH, ORD | LL, EWR | LH, JFK |
| 20 | 4 | LH, ORD | HL, EWR | HH, JFK | LL, LAX |
| 21 | 5 | LL, LAX | LH, EWR | HL, JFK | HH, ORD |
| 22 | 6 | LH, EWR | HL, JFK | HH, ORD | LL, LAX |
| 23 | 7 | HL, JFK | HH, ORD | LL, LAX | LH, EWR |
| 24 | 8 | HH, ORD | LH, LAX | HL, JFK | LL, EWR |
| 25 | 1 | LL, EWR | LH, JFK | HL, LAX | HH, ORD |
| 26 | 2 | HH, JFK | LL, LAX | LH, ORD | HL, EWR |
| 30 | 3 | HL, LAX | HH, ORD | LL, EWR | LH, JFK |

*Note*. LH – Low visual density, high ID, HH – High visual density, high ID, HL- high visual density, low ID, LL – low visual density, low ID. EWR – Newark International Airport. ORD – Chicago O'Hare International Airport. LAX – Los Angeles International Airport. JFK – John F. Kennedy International Airport.

# Appendix J

## Raw Data

| | Low VD, Low ID | | | Low VD, High ID | | | High VD, Low ID | | | High VD, High ID | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | SA | Workload | SP[1] | SA | Workload | SP | SA | Workload | SP | SA | Workload | SP |
| 1 | 10 | 75.32 | 3 | 14 | 34.17 | 4 | 8 | 61.26 | 14 | 12 | 28.33 | 9 |
| 2 | 8 | 55.83 | 5 | 14 | 20.83 | 3 | 12 | 65.36 | 16 | 10 | 39.17 | 6 |
| 3 | 9 | 56.67 | 6 | 13 | 36.67 | 3 | 6 | 54.17 | 18 | 10 | 62.50 | 4 |
| 4 | 9 | 45.36 | 7 | 15 | 43.33 | 4 | 9 | 38.33 | 10 | 18 | 34.17 | 11 |
| 5 | 11 | 62.50 | 10 | 14 | 45.23 | 5 | 7 | 73.33 | 15 | 13 | 67.50 | 5 |
| 6 | 7 | 94.17 | 8 | 11 | 55.23 | 4 | 4 | 56.26 | 10 | 7 | 65.00 | 6 |
| 7 | 7 | 41.67 | 8 | 9 | 63.33 | 3 | 6 | 74.17 | 18 | 10 | 65.00 | 5 |
| 8 | 6 | 45.00 | 9 | 8 | 52.89 | 5 | 7 | 43.33 | 12 | 9 | 41.67 | 9 |
| 9 | 7 | 78.33 | 9 | 13 | 65.83 | 4 | 10 | 71.67 | 10 | 10 | 35.00 | 4 |
| 10 | 7 | 46.67 | 6.4 | 12 | 40.83 | 4 | 5 | 58.11 | 13.48 | 6 | 54.17 | 6.26 |
| 11 | 5 | 66.67 | 6.4 | 9 | 61.36 | 4 | 6 | 73.33 | 13.48 | 5 | 71.67 | 4 |
| 12 | 8 | 56.32 | 7 | 10 | 54.17 | 4 | 11 | 51.67 | 12 | 14 | 37.50 | 6 |
| 13 | 9 | 59.17 | 5 | 10 | 57.35 | 11 | 8 | 75.83 | 17 | 8 | 78.33 | 5 |
| 14 | 8 | 74.17 | 9 | 14 | 25.83 | 4 | 9 | 33.33 | 14 | 11 | 67.50 | 8 |
| 15 | 6 | 58.33 | 4 | 14 | 54.17 | 6 | 3 | 75.00 | 17 | 10 | 78.33 | 3 |
| 16 | 8 | 72.50 | 7 | 10 | 68.33 | 3 | 4 | 60.00 | 13.48 | 9 | 55.83 | 6 |
| 17 | 6 | 45.00 | 3 | 13 | 44.17 | 5 | 6 | 15.00 | 8 | 8 | 29.17 | 4 |
| 18 | 11 | 55.00 | 8 | 11 | 44.17 | 4 | 11 | 49.17 | 13 | 11 | 51.67 | 8 |
| 19 | 9 | 36.67 | 7 | 13 | 51.67 | 6 | 4 | 98.33 | 10 | 9 | 39.33 | 7 |

| | Low VD, Low ID | | | Low VD, High ID | | | High VD, Low ID | | | High VD, Low ID | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | SA | Workload | SP | SA | Workload | SP | SA | Workload | SP | SA | Workload | SP |
| 20 | 6 | 50.00 | 6.4 | 12 | 48.33 | 7 | 6 | 35.83 | 18 | 10 | 37.50 | 7 |
| 21 | 9 | 57.50 | 5 | 12 | 57.50 | 8 | 9 | 65.00 | 14 | 10 | 50.70 | 4 |
| 22 | 5 | 44.17 | 6.4 | 10 | 28.33 | 3 | 9 | 21.67 | 6 | 11 | 30.83 | 8 |
| 23 | 6 | 40.83 | 5 | 9 | 45.00 | 5 | 9 | 48.33 | 16 | 9 | 45.00 | 8 |
| 24 | 9.04 | 73.33 | 3 | 10 | 54.17 | 4 | 8 | 51.67 | 15 | 10.09 | 50.83 | 7 |
| 25 | 9 | 60.00 | 8 | 11 | 58.33 | 6 | 5 | 85.00 | 14 | 15 | 94.17 | 6 |
| 26 | 9 | 100.00 | 9 | 13 | 91.67 | 4.76 | 13 | 74.17 | 13 | 14 | 98.33 | 8 |

*Note*. [1] Search performance.